11-12-2018

# Speciation and hybridization in Jamaican-endemic Streamertail Hummingbirds (Trochilus polytmus and T. scitulus)

Caroline D. Judy
*Louisiana State University and Agricultural and Mechanical College*, caroline.duffie@gmail.com

# SPECIATION AND HYBRIDIZATION IN JAMAICAN-ENDEMIC STREAMERTAIL HUMMINGBIRDS (*TROCHILUS POLYTMUS* AND *T. SCITULUS*)

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences
and
Museum of Natural Science

Caroline D. Judy
B.Sc. University of Georgia, 2002
M.Sc. University of Missouri-St. Louis, 2007
December 2018

John O’Neill, 1976

To Dan and our Rose.

## ACKNOWLEDGEMENTS

I have many people to thank for helping me complete this dissertation. First, I thank my co-advisors, Dr. Robb T. Brumfield and Dr. Gary R. Graves, who have heavily invested in my intellectual development as well as this body of research. I would also like to thank my other committee members, Dr. Jessica Eberhard, Dr. Chris Austin, and Dr. Tom Tully, for their guidance. Jessica, your steady mentorship since my first year in the program has been a big source of encouragement. Chris, I appreciate your willingness to serve so late in the game! I have a fabulous committee.

It has been a privilege to work in Jamaica. I thank Dr. Susan Koenig and the late Mr. Mike Schwartz (1947–2018) of Windsor Research Centre for allowing me to collect streamertails on Windsor property, and for their hospitality to me on many other occasions. I thank Ms. Linnette Wilkes of the Bowden Pen Farmers Association for supplying me with trustworthy guides, especially Mr. Eric McCurbin, who worked with me through multiple seasons. I thank the many other field assistants who made my work possible. Matthew L. Brady assisted with the streamertail collections. His skill as a preparator was invaluable for the project. I thank Catherine Levey, who supplied me with botanical guides and other resources pertaining to natural history in Jamaica. I thank Mr. Ricardo Miller, Ms. Monique Curtis, and Ms. Andrea Donaldson (National Environment and Planning Agency) for assisting with permits. Ms. Ellen Paul (Ornithological Council) facilitated communications between U.S. and Jamaican agencies to gain final import approval for specimen import. I am very thankful for her efforts.

I am grateful to Ms. Tammie Jackson at the LSU Museum of Natural Science for her skillful administrative support over the last eight years. She is just wonderful. Ms.

Valerie Derouen (outreach coordinator) has also been very helpful. I thank Donna L. Dittman for her excellent help with import permits. The staff of the Department of Biological Sciences has also been very supportive, especially Ms. Chimene Williams.

I would like to thank the former graduate students of the Brumfield lab for their critical feedback as I worked to develop this project: Dr. Andres Cuervo, Dr. Gustavo Bravo, Dr. James Maley, Dr. Sarah Hird, Dr. Michael Harvey, and Dr. Glenn Seeholzer. I also would like to thank the current students: Mr. Matthew Brady, Ms. Glaucia Del-Rio, Ms. Anna Hiller, Mr. Oscar Johnson, Mr. Rafael Marcondes, Mr. Andre Moncrieff, Mr. Marco Rego, and Ms. Jessie Salter. Many faculty members have also provided helpful discussion: Dr. James Van Remsen, Dr. Fred Sheldon, Dr. Prosanta Chakrabarty, Dr. Meheshi Dassanayake, Dr. Jeremy Brown, Dr. Michael Hellberg, and Dr. Bryan Carstens (now at The Ohio State University). I thank other friends, formerly or currently associated with the LSU Museum of Natural Science and Biological Sciences: Mr. Dan Lane, Dr. Eric Rittmeyer, Dr. Noah Reid, Dr. Melissa DeBiasse, Dr. Richard Gibbons, Dr. Verity Mathis, Dr. Caleb McMahan, Dr. Brian Smith, Dr. Ryan Terrill, Dr. Jessica Oswald, Dr. Cathy Newman, Dr. Bill Ludt, Ms. Vivien Chua, Ms. Clare Brown, Mr. Ryan Burner, Ms. Genevieve Mount, and Mr. Zach Rodriguez.

I would like to thank the entire Bird Division at the National Museum of Natural History (NMNH) where I have been a resident fellow since 2013. Collections staff members Mr. Brian Schmidt, Ms. Christina Gebhard, Mr. Jacob Saucier, and Mr. Christopher Milensky have been excellent resources. Mr. Brian Schmidt assisted with Institutional Animal Care and Use Committee (IACUC) paperwork and protocols, and his expertise in Jamaica has been a big help. I thank Dr. Helen James, Curator in Charge,

who has been very supportive of me. I also thank the Laboratory of Analytic Biology (LAB) director, Mr. Lee Weigt, and manager, Mr. Jeffrey Hunt, for facilitating my project in the lab. I am very grateful for the tremendous bioinformatics support offered by Mr. Matthew Kweskin, Dr. Vanessa González, Dr. Rebecca Dikow, and Mr. Mike Trizna. They are an amazing team. Finally, I thank many colleagues formerly or currently associated with the NMNH community for helpful discussions: Dr. Mike Braun, Dr. Robert Fleischer, Dr. Mirian Tsuchiya, Dr. Noor White, Dr. Haw Chuan Lim, Dr. Rayna Bell, Dr. Pete Hosner, Dr. Adela Roa-Varon, Dr. Niamh Redmond, Dr. Amy Driskell, Dr. Teresa Feo, Ms. Catherine Early, and Ms. Caitlin Wells. The training course in R programming language taught by Dr. Gene Hunt and his generous help over email has enabled me to achieve R literacy, which is invaluable skill set for my career. In sum, I am appreciative of the NMNH community for their friendship and investment in my research.

I am deeply grateful for my husband, Dan Judy, and daughter, Rosie. No one deserves more thanks than Dan. His unwavering love and support have sustained me through the many ups and downs of this project. Thank you, Dan, for believing in me, and making my goals your own! Rosie has been our little sunbeam through it all, and I am grateful for the constant joy she has given us over the last 2½ years. Finishing a PhD with a beautiful family around me is just about the most wonderful thing I could imagine. I would also like to thank my parents, Carol and Traywick, and my sisters, Amanda and Lauren, who have always supported my endeavors. My parents-in-law, Patsy and Richard, and sister-in-law, Laura, have been constant encouragers, which is their gift. Pursuits give life purpose, but family makes it special. I am thankful to God for the

opportunity to study at Louisiana State University and for giving me the family that has supported me along the way.

**FINANCIAL SUPPORT**

My field work was supported with a James Bond Fund of the Smithsonian Institution through an award to Gary R. Graves, the LSU Museum of Natural History, and the American Museum of Natural History (Chapman Grant). Robb T. Brumfield paid for all laboratory expenses with general laboratory funds and an NSF grant.

# TABLE OF CONTENTS

**ABSTRACT**

Streamertails hummingbirds (*Trochilus polytmus* and *T. scitulus*) are recently diverged sister taxa that appear to have speciated *in situ* on the island of Jamaica. They are distinguished by male bill color, a secondary sexual trait that is coral red in *T. polytmus* and jet black in *T. scitulus*. They hybridize in a narrow zone where their ranges meet in eastern Jamaica. In Chapter 2, I performed a formal population survey of *T. scitulus* to determine the size of the population, which was unknown. I determined that the total population contains well over 100,000 individuals despite its limited geographic range. In Chapter 3, I build on previous studies to identify divergent morphological and genetic traits. Additionally, I use geographic cline models to determine the center and widths of individual clines and make inferences about the relative strength of selection acting on each trait. The clines for male bill color (2.2 km) and bill width (13.9 km) were narrow relative to neutral expectations and centered on the Rio Grande Valley. Female values were slightly wider. Consistent with expectations for recently diverged species, I detected little to no neutral genomic differentiation using six microsatellites, but a Radseq dataset containing 6,451 single nucleotide polymorphisms (SNPs) generated by a genotyping-by-sequencing protocol showed modest genetic differentiation between species. A narrow subset of SNPs (n = 23) with high loadings in a discriminant function analysis may be physically or epistatically linked to the divergent morphological traits. A structure analysis based on these discriminant SNPs shows a range of admixture assignments in the hybrid zones and strong differences in assignment between the parental species. A geographic cline analysis based on admixture assignments revealed a narrow cline (6.6 km) also centered on the Rio Grande. Finally, I performed a transcriptomics study to

examine baseline differences in gene expression. While gene expression profiles were extremely similar, four genes showed significant gene expression differences. One gene, *BLOC-1S1*, is a ubiquitously expressed gene that is associated with pigmentation disorders in mice and humans. Its role in avian pigmentation is not well characterized, but is one gene that warrants further investigation as a candidate gene underpinning male bill color. I also looked at sequence divergence across the assembled transcripts. This panel of SNPs showed low divergence and little evidence for positive selection, indicating that divergent genes may be few and potentially located in regulatory regions not captured here.

# CHAPTER 1.

# INTRODUCTION

Understanding the drivers of speciation in nature is an overarching goal of evolutionary biology. Molecular genetic studies have contributed to the view that speciation is a continuum (Powell et al. 2013; Seehausen et al. 2014). In this view, parapatric races and their hybrid zones may be seen as representative stages in the divergence process (Hewitt 1988). Such 'races' of incipient taxa are important model systems for examining divergent phenotypes and their role in speciation before reproductive isolation has evolved (Shaw and Mullen 2011; Nosil and Feder 2013; Powell et al. 2013; Seehausen et al. 2014). Because early-stage divergence can transition rapidly (Gavrilets 2003), catching races in this early snapshot of speciation can be challenging.

I use the recently diverged streamertail hummingbirds (Trochilidae: *Trochilus*) as a study system to examine hybridization and speciation. These Jamaican endemics offer a rare glimpse of incipient speciation to occur *in situ* within the confines of a small oceanic island. They are distinguished primarily by male bill color (Fig. 1.1), which is bright red in the red-billed streamertail (*Trochilus polytmus*) and jet black in the black-billed streamertail (*T. scitulus*). They are considered subspecies by the American Ornithologists Society (Chesser et al. 2018). However, Karl Schuchmann treats them as full species based on noted differences in courtship and song (Schuchmann 1978, 1980).

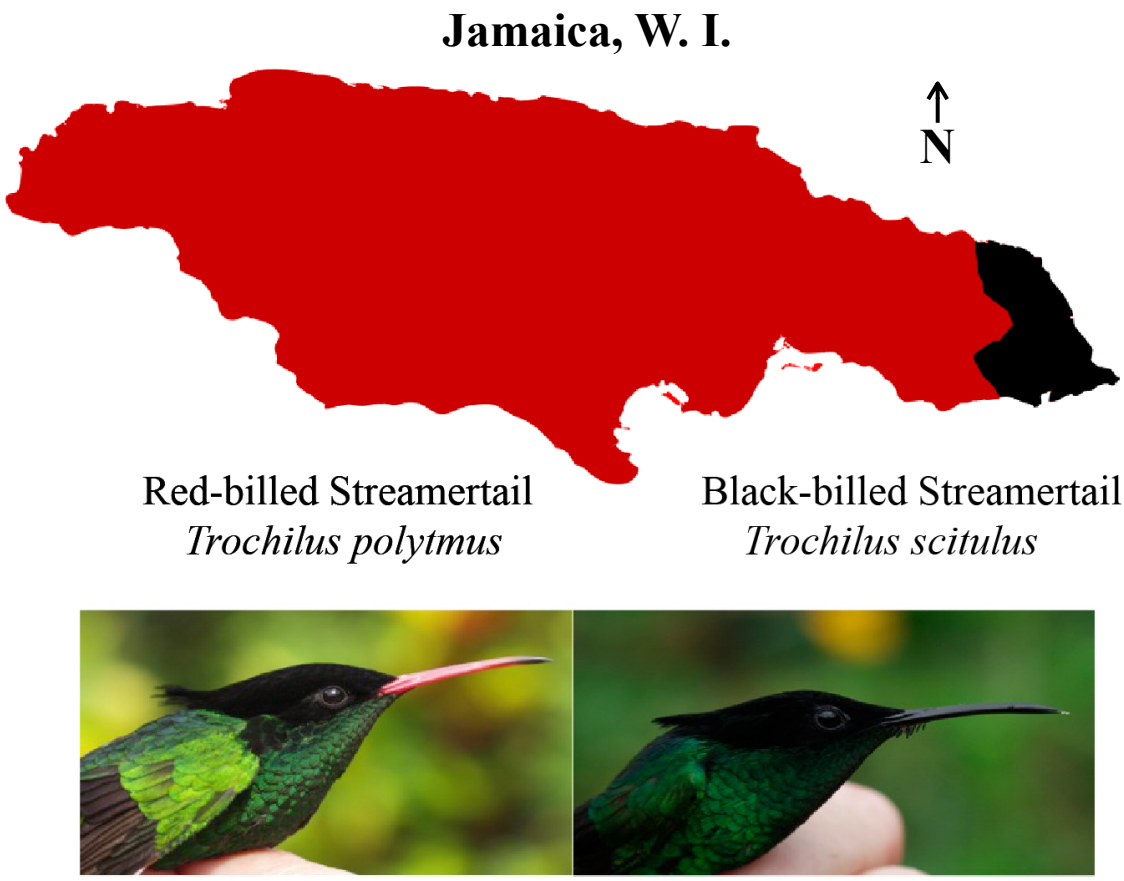



Fig. 1.1. Schema depicting the approximate ranges of *Trochilus polytmus* and *T. scitulus* on Jamaica, West Indies. Photographs are of *T. polytmus* (left) and *T. scitulus* (right). Photo credit: Gary R. Graves.

In Chapter 2, I determine the previously unknown population size and density of the black-billed streamertail. The black-billed streamertail and the red-billed streamertail form parapatric distributions on the island, but the red-billed streamertail is much more widespread. They form a narrow hybrid zone where their ranges meet in eastern Jamaica. The IUCN Red List classifies the black-billed streamertail as Least Concern (Conservation International 2012) based on observations that it appears common where it occurs. However, a formal survey is warranted because its entire geographic range is less than 600 $km^2$. This small range makes the species inherently more vulnerable to

extinction. I conducted a point count survey (n = 530) and used distance-based approaches to estimate population density within the study range. Robust density and abundance estimates are useful for setting conservation and management strategies and inform hybrid zone dynamics.

Hybrid zones are natural laboratories to examine the characters and processes involved in divergence and the evolution of barriers to gene flow (Hewitt 1988; Harrison 1990). Barriers to gene flow are broadly classified as prezygotic or postzygotic (Ernst Mayr 1942; Dobzhansky 1982). In birds, the impressive diversity of display traits used to attract mates has inspired the hypothesis that prezygotic mechanisms may potentially have a bigger role (Uy et al. 2018). In this view, diversifying sexual selection may figure prominently in avian speciation (West-Eberhard 1983), which is supported by the observation that even distantly related taxa can hybridize successfully though close relatives can coexist sympatrically (reviewed in Uy et al. 2018).

In Chapter 3, I examined the morphological and genetic structure of the hybrid zone between streamertail hummingbirds. I characterized two morphological traits previously shown to be divergent: bill color and bill width (Brewster and Bangs 1901; Gill et al. 1973; Graves 2009a, 2015). I also measured variation in skeletal features, which I used as correlates of body size, following methods in (Graves 2009b). To examine genetic traits, I characterized six microsatellites and a 6,451 SNPs from a genotyping by sequencing (GBS) dataset. I used a combination of clustering and multivariate approaches to identify informative loci. I fit cline models to morphological and genetic traits that showed clinal variation, and use the estimated cline widths to make inferences about selection.

In Chapter 4, I use a transcriptomics approach to study the genetic and regulatory basis for divergence between streamertails. RNA sequencing (RNA-seq) is a cost-effective technology that has enabled transcriptome profiling in nonmodel organisms at an unprecedented rate (Toews et al. 2015; Jax et al. 2018). I assembled a transcriptome based on seven libraries of streamertail hummingbirds and used Blast2Go to annotate it. I performed an analysis of differential gene expression and used alignments of the raw sequencing reads to the *de novo* assembly to call genome-wide SNPs. I explored the function of the differentially expressed genes and looked for signatures of positive selection in the SNPs using an $F_{ST}$ outlier method.

The most obvious morphological trait difference between streamertails is male bill color, which is coral red and narrowly tipped in black (partially melanized) in the red-billed streamertail (*Trochilus polytmus*), and jet black (fully melanized) in the black-billed streamertail (*T. scitulus*). In field-based collections, the red coloration of *T. polytmus* fades rapidly post mortem, suggesting that the bill appears red due to the presence of highly vascularized tissues rather than a red pigment such as a carotenoid or phyomelanin (Graves 2009a, 2015). For this reason, I hypothesize that bill color is controlled by melanin-pathway genes, which have been well studied in other vertebrates, especially humans and laboratory strains of mice (Hoekstra 2006), and to a lesser extent in birds (Poelstra et al. 2013). I used this rich knowledge base to create a custom dataset of known pigmentation genes to further annotate my transcriptome assembly, and for use in future enrichment and pathway studies in conjunction with the panel of discovered RNA-seq SNPs.

In sum, I provide the first formal survey of population size and density for the black-billed streamertail (Chapter 2), which is useful as a benchmark study of population size going forward. In Chapter 3, I performed the first comprehensive analysis of the streamertail hybrid zone between incipient species of streamertail hummingbirds in the genus *Trochilus* and make inferences regarding the role of selection to promote divergence. Finally, in Chapter 4, I assembled a *Trochilus* transcriptome and examine differences in expression profiles and sequence divergence between these recent species. Collectively, this body of work addresses the morphological and genetic changes that are emerging in recently diverged lineages to shed light on what might be driving speciation within the confines of a small oceanic island.

**REFERENCES**


Bird Life International, 2012. *Trochilus scitulus*. The IUCN red list of threatened species. Version 2014.1. Downloaded on 15 July 2014.

Brewster, W., and O. Bangs. 1901. On an overlooked species of *Aithurus*. Proc. N. Engl. Zoöl. Club 2:47–50.

Charles Darwin. 1871. The descent of man, and selection in relation to sex. Princeton University Press, Princeton, New Jersey.

Chesser, R. T., K. J. Burns, C. Cicero, J. L. Dunn, A. W. Kratter, I. J. Lovette, P. C. Rasmussen, J. V. Remsen, Jr., D. F. Stotz, B. M. Winger, K. Winker 2018. Check-list of north American birds (online). American Ornithological Society. http://checklist.aou.org/taxa

Dobzhansky, T. 1982. Genetics and the origin of species. Columbia University Press.

Ernst Mayr. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press, New York.

Galván, I., and F. Solano. 2016. Bird integumentary melanins: biosynthesis, forms, function and evolution. Int. J. Mol. Sci. 17.

Gavrilets, S. 2003. Perspective: models of speciation: what have we learned in 40 years? Evolution 57:2197–2215.

Gill, F. B., F. J. Stokes, and C. Stokes. 1973. Contact zones and hybridization in the Jamaican hummingbird, *Trochilus polytmus* (L.). The Condor 75:170–176.

Graves, G. R. 2015. A primer on the hybrid zone of Jamaican streamertail hummingbirds (Trochilidae: *Trochilus*). Proc. Biol. Soc. Wash. 128:111–124.

Graves, G. R. 2009a. Ontogeny of bill color in male streamertail hummingbirds (*Trochilus*). J. Caribb. Ornithol. 44–47.

Graves, G. R. 2009b. Skeletal correlates of body weight in the black-billed streamertail (*Trochilus scitulus*) of Jamaica. Caribb. J. Sci. 45:69–72.

Harrison, R. G. 1990. Hybrid zones: windows on evolutionary process. Oxf. Surv. Evol. Biol. 7:69–128.

Hewitt, G. M. 1988. Hybrid zones-natural laboratories for evolutionary studies. Trends Ecol. Evol. 3:158–167.

Hoekstra, H. E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97:222–234.

Jax, E., M. Wink, and R. H. S. Kraus. 2018. Avian transcriptomics: opportunities and challenges. J. Ornithol. 159:599–629.

Nosil, P., and J. L. Feder. 2013. Genome evolution and speciation: toward quantitative descriptions of pattern and process. Evolution 67:2461–2467.

Poelstra, J. W., H. Ellegren, and J. B. W. Wolf. 2013. An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. Heredity 111:467–473.

Powell, T. H. Q., G. R. Hood, M. O. Murphy, J. S. Heilveil, S. H. Berlocher, P. Nosil, and J. L. Feder. 2013. Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Diptera: Tephritidae). Evolution 67:2561–2576.

Schuchmann, K.-L. 1978. Allopatrische artbildung dei der kolibrigattung *Trochilus*. Ardea 66:156–172.

Schuchmann, K.-L. 1980. Die Jamaika kolibris *Trochilus polytmus* und *Trochilus scitulus*. Biotropic-Verlag, Frankfurt am Main.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–192.

Shaw, K. L., and S. P. Mullen. 2011. Genes versus phenotypes in the study of speciation. Genetica 139:649–661.

Toews, D. P. L., L. Campagna, S. A. Taylor, C. N. Balakrishnan, D. T. Baldassarre, P. E. Deane-Coe, M. G. Harvey, D. M. Hooper, D. E. Irwin, C. D. Judy, N. A. Mason, J. E. McCormack, K. G. McCracken, C. H. Oliveros, R. J. Safran, E. S. C. Scordato, K. F. Stryjewski, A. Tigano, J. A. C. Uy, and B. M. Winger. 2015. Genomic approaches to understanding population divergence and speciation in birds. The Auk 133:13–30.

Uy, J. A. C., D. E. Irwin, and M. S. Webster. 2018. Behavioral isolation and incipient speciation in birds. Annu. Rev. Ecol. Evol. Syst., doi: 10.1146/annurev-ecolsys-110617-062646.

West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. Q. Rev. Biol. 58:155–183.

## CHAPTER 2.

## DENSITY AND ABUNDANCE OF THE BLACK-BILLED STREAMERTAIL (*TROCHILUS SCITULUS*) IN EASTERN JAMAICA

### INTRODUCTION

The black-billed streamertail (*Trochilus scitulus*) is a Jamaican endemic hummingbird that is geographically restricted to the extreme eastern end of the island (Brewster and Bangs 1901, Gill et al. 1973, Graves 2015). The black-billed streamertail is replaced to the west by its congener, the red-billed streamertail (*Trochilus polytmus*, Linnaeus 1758), with which it hybridizes (Gill et al. 1973; Graves 2015). The males of both species possess the velvet black crests, emerald gorgets, and elaborate tail plumes from which their common name is derived (Fig. 2.1), but are distinguished by bill color, which is jet black in the black-billed streamertail, and coral red in the red-billed streamertail. Despite its much smaller geographic range, the black-billed streamertail is currently classified by the IUCN as 'Least Concern' ver 3.1 (Birdlife International 2012) due to the fact that it appears common where it occurs. However, a formal population survey is needed to more rigorously assess its population status.

The black-billed streamertail range in the John Crow and Blue Mountains is composed primarily of wet montane forest and surrounding coastal lowlands. The north-facing slopes receive the most rain on the island. Here, the black-billed streamertail occurs east of the Rio Grande Valley. Along Jamaica's southern coast, it is found east of the Morant River Valley. The red-billed streamertail is widely distributed across most of the rest of the island. Where their parapatric ranges meet in the Rio Grande Valley they form a narrow zone of hybridization (Gill et al. 1973, Graves 2015). These sister taxa

Fig. 2.1. Mature black-billed streamertail (*Trochilus scitulus*) near Bath in eastern Jamaica. Photo by Caroline Judy.

may represent a rare example of *in situ* speciation to occur on a small island (Coyne and Price 2000); Jamaica is less than 11,000 km$^2$. The distributional limits of the two species and the geographic boundaries of their hybrid zone are well characterized (Gill et al. 1973). However, the evolutionary mechanisms that maintain their species-level distinctiveness in the face of ongoing hybridization are unknown and currently under investigation (Graves 2009a, b, 2015; Lance et al. 2009; McCormack et al. 2012).

Both the male and the female black-billed streamertail produce call notes (Schuchmann 1977). The black-billed streamertail call is a simple repeated high-pitched note that can be heard up to 200 m away, depending on habitat features. The typical male

song is a squeaky, trilling vocalization (Schuchmann 1977, 1980). In addition, the fluttering wing feathers of adult males produce a shrill whirring noise (Gosse 1847; Clark 2008). This fluttering is detectable at close range (i.e. within 15 m). Little is known about female song, and females do not produce the whirring noise in flight.

Schuchmann (1999) speculates that there are approximately 12 pairs/km$^2$ of red-billed streamertail in the Blue Mountains and 3-6 pairs/km$^2$ of black-billed streamertail in the vicinity of Port Antonio (Portland Parish) and Bath (St. Thomas Parish), though the year(s) and season(s) on which these estimates are based are unspecified. Wunderle et al. (1992) report that prior to Hurricane Gilbert in 1988, red-billed streamertail densities were 0.77 individuals per census point in montane cloud forest habitats, with lower densities in other habitats and at lower elevations. While Schuchmann's observations suggest that black-billed streamertail density may be lower than that of red-billed streamertail, the lack of current population information for either species limits interspecific comparisons. The smaller range size and potentially lower density of the black-billed streamertail may make it more susceptible to population declines related to habitat perturbation. Here, I conduct a point count survey and use conventional distance sampling methods (Buckland et al. 2001) to provide a rigorous assessment of the black-billed streamertail population density and global abundance.

**METHODS**

The point count survey protocol was field tested in January 2014 on a red-billed streamertail population in Trelawny Parish. I then conducted a survey of the black-billed streamertail population over 18 days (5-23 February 2014) in Portland and St. Thomas Parishes during daylight hours between 06:15 and 18:00h. The rough terrain and limited

access to large regions within the range of the black-billed streamertail prevented the use of a randomized survey design. I conducted the point count survey along roads, community footpaths, and forestry trails. Roughly 90% of the survey was conducted via a four-wheel vehicle. I traveled on foot along community footpaths in a few areas where there were no larger roads. I spaced the survey points at least 200 m apart in order to minimize counting the same bird at multiple consecutive census points.

I surveyed primary and secondary montane rain forest, gardens, orchards, landscaped residential areas, mixed agricultural areas such as shaded coffee groves, forested perimeters of banana and coconut plantations, dasheen fields, pasturelands, and patchworks of small agricultural plots near homesteads. A point was considered suitable if it supported as least 10% forest cover, loosely defined, within a 15 m radius of the station. Additionally, points had to be located within 10 m of a tree that was at least 10 m high. These criteria helped to exclude poor habitat areas such as large monocultures of sugar cane and other crops, open pastures, and fallow fields. I also avoided surveying near town centers, church services, schools in session, or noisy homesteads where human activities hindered the survey effort.

After arriving to a point, I recorded the date, time of day, geographic coordinates, and elevation using a Garmin® Vista HCx GPS. I made notes on the weather conditions and habitat features, including visible nectar sources. After these data were recorded, at least two observers (the author and one other) scanned the area for black-billed streamertail using Nikon Monarch 8x42 binoculars during a three-minute counting period. The horizontal distance to each detected individual was measured or estimated using a Simmons® LRF 600 rangefinder or a Garmin® Vista HCx GPS and recorded in a

field notebook. After the counting window, additional notes on nectar sources were also recorded, as well as presence of other nectar feeding birds. Finally, digital photographs were taken to further document habitat features.

The black-billed streamertail was detected visually and aurally. For each detected streamertail, I noted whether the detection was fluttering, calling, singing, or visual. In regions that are geographically within or near the contact zone, I spent additional effort to identify each individual as a black-billed streamertail, red-billed streamertail, or putative hybrid. Details on sex, molt, and behavior were also recorded when possible. These data were used collectively to determine the minimum number of black-billed streamertail individuals at each point.

The survey area encompasses most of the known geographic range for the black-billed streamertail (Fig. 2.2). I constructed a minimum convex polygon (MCP) around all survey points using the minimum bounding geometry tool in ArcGIS software v. 10.2.

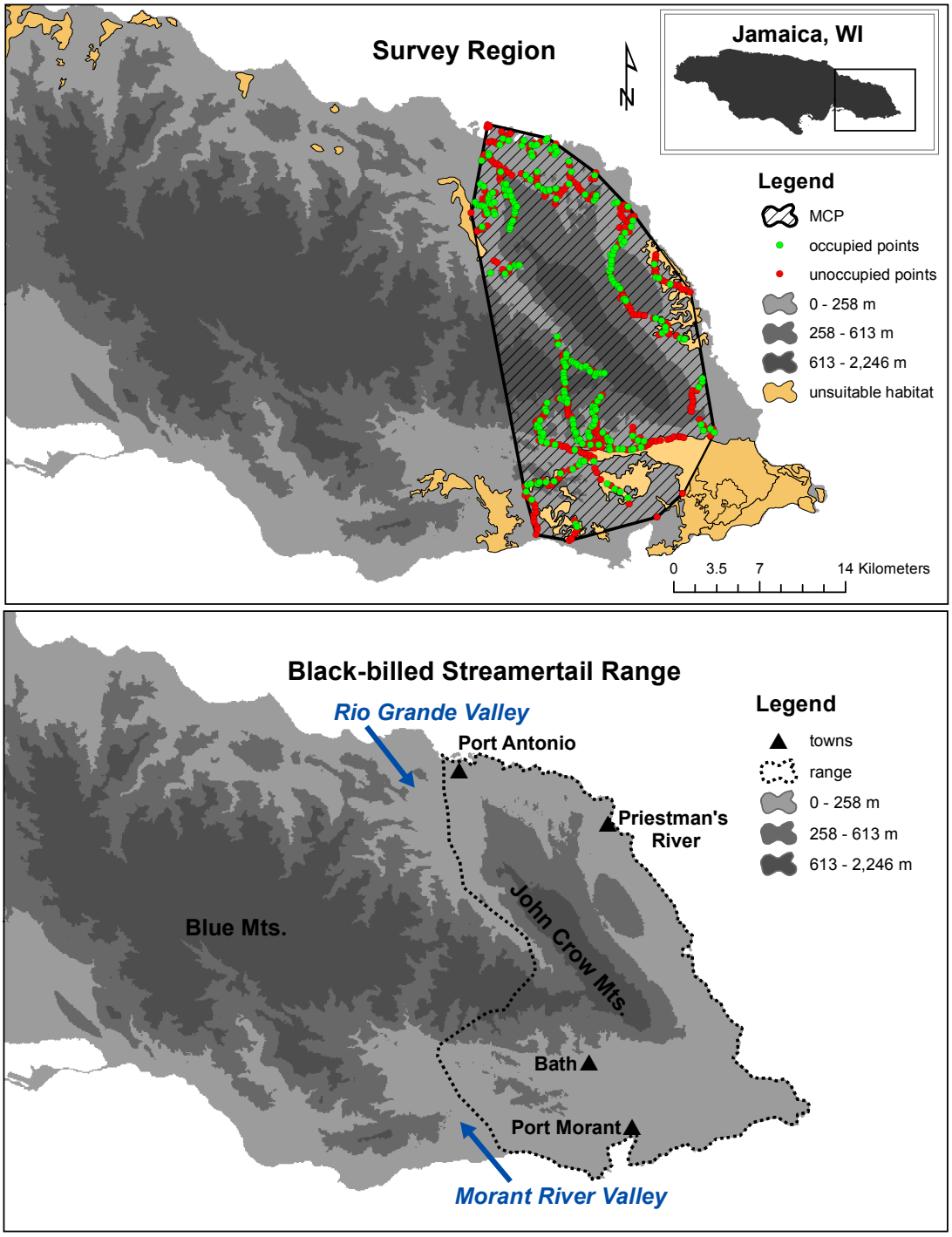



Fig. 2.2. Top: Survey region. The shaded polygon represents the minimum convex polygon (MCP). Bottom: Black-billed streamertail range. The exact area of transition from the red-billed to black-billed streamertail in the Blue Mountains is uncertain.

The total area of the resulting MCP is 453 $km^2$. I assessed the habitat suitability within the MCP using the LU1984 coverage layer available from the Forestry Department (http://www.forestry.gov.jm/maps_data_page.html). I designated regions as "unsuitable" that are described in the LU1984 coverage layer as "tree crops, shrub crops, sugar cane, and banana." While small agricultural areas and the tree-lined edges of large agricultural fields are suitable for the black-billed streamertail, the deforested areas are not. Removing these areas from the MCP yields an adjusted area of 422 $km^2$. I also used 2012 Google Earth Pro satellite images of Portland and St. Thomas Parishes to identify additional areas of deforestation. Visual inspection of Google satellite images revealed that an additional 15 - 20% of the MCP may be deforested or otherwise heavily impacted. Therefore, I conservatively set the size of the survey area to be 329 $km^2$, or 80% of the adjusted land area within the MCP.

*Distance Sampling*

Distance sampling allows rigorous estimation of density in the face of variability in detection (Buckland et al. 2001). A key assumption is that points are located at random with respect to the birds, such that changes in the number of birds observed with increasing distance from the point can be interpreted as changes in detectability, rather than density. More specifically, distance sampling uses the distribution of the observed distances to estimate a detection function g(y), the probability of detecting a bird at distance y. This function can then be used to estimate the average probability of detecting a bird within distance $w$ of the point, denoted $P_a$. Given an estimate of $P_a$, bird density can be estimated as

$$D = n/aP_a$$

where $n$ is the number of birds detected and $a$ is the size of the covered region. Finally, abundance can be estimated by extrapolation of the estimated density within the covered region to the larger study area.

I evaluated the fit of the uniform, half-normal, and hazard-rate detection models with quantile-quantile plots and goodness-of-fit tests (Buckland et al. 2001) to six data filters corresponding to three data types (ungrouped, grouped using five equal cutpoints away from favored numbers, grouped using 8 equal cutpoints), and two truncation schemes ($w$ = 40 meters, no truncation). Model selection was made using the Akaike's Information Criterion (AIC) (Akaike 1973), and precision was estimated analytically (see details in Fewster et al. 2009; Thomas et al. 2010). All analyses were implemented in the Distance 7.2 software (Thomas et al. 2010).

*Variable Circular Plot method*

I used the variable circular plot method of (Reynolds et al. 1980) to provide an independent calculation of density and abundance for the black-billed streamertail population. I summarized the detections by their horizontal distances in 5 m bands around the point, i.e. at 5, 10, 15, 20 m, etc. I then calculated the density of detections in each band by dividing the number of detections by the area of that band. The inflection point, or the distance at which detection densities decline more than 50%, serves as a guideline for drawing the effective census radius, beyond which observations are discarded (additional details in Reynolds et al. 1980). The calculation for density is based on only observations that fall within the effective census radius. Finally, I estimated abundance as

the product of density, occupancy rate (proportion of points that were occupied vs. unoccupied), and the size of the survey region.

## RESULTS

### *Distance Sampling*

I detected a total of 287 individuals at 530 survey points (Fig. 2.2). Aural detections (calling, singing, and fluttering; n = 159) constituted 55% of all detections (Fig. 2.3A). Fluttering had the closest mean distance from the point (6.60 m), and calling had the farthest (31.4 m). The maximum detection distance was 80 m, considerably greater than the second largest distance, 60 m. As is common for point count surveys, there was a long tail of distribution; the median distance was 15 m.

Visual inspection of the data revealed a signal for an excess of observations at zero distance from the point, or spiking, which indicates rounding to zero. There was also a strong pattern of favored numbers (heaping), and appearance of over-dispersion (Fig. 2.3B), suggesting the data should be grouped into intervals for reliable analysis. Truncation of distances to 40 m removed 21 or 7.3% of observations from the dataset. Truncation improved reliability of the goodness-of-fit tests by increasing the expected values in the largest distance intervals, though it reduced precision (Table 2.1 & Table 2.2).

Table 2.1. Summary of AIC values for two truncation values, $w = 40$, and w = largest observation. The data were analyzed ungrouped and with two different groupings: five intervals of equal widths with cutpoints away from favored values, and eight intervals of equal widths. For each data type, an asterisk indicates which model had the smallest AIC score. AIC scores for failed models are indicated by "n.a."

| | | $w = 40$ | | | $w$ = largest observation | | |
|---|---|---|---|---|---|---|---|
| | | No. of parameters | | | No. of parameters | | |
| Data type | Model (Key + adjustment) | Key | Adjustments | AIC | Key | Adjustments | AIC |
| Ungrouped | Half normal + cosine | 1 | 3 | 3,503.8 | 1 | 4 | 3,831.8 |
| | Half normal + Hermite | 1 | 0 | n.a. | 1 | 0 | n.a. |
| | Uniform + cosine* | 0 | 4 | 3,300.6 | 0 | 5 | 3,660.1 |
| | Hazard rate + cosine | 2 | 0 | 3,311.3 | 2 | 0 | 3,738.1 |
| Grouped | Half normal + cosine* | 1 | 2 | 820.3 | 1 | 2 | 644.6 |
| (5 equal) | Half normal + Hermite | 1 | 0 | 896.4 | 1 | 0 | 700.1 |
| | Uniform + cosine | 0 | 4 | 823.2 | 0 | 2 | 789.7 |
| | Hazard rate + cosine | 2 | 0 | 827.2 | 2 | 0 | 650.8 |
| Grouped | Half normal + cosine | 1 | 4 | 1,048.2 | 1 | 4 | 848.1 |
| (8 equal) | Half normal + Hermite | 1 | 0 | 1,204.6 | 1 | 0 | 934.9 |
| | Uniform + cosine | 0 | 5 | 1,061.7 | 0 | 5 | 888.5 |
| | Hazard rate + cosine* | 2 | 0 | 1,038.4 | 2 | 0 | 864.8 |

The data filter with 5 equal cutpoints and truncation at 40 meters had cutpoints away from favored numbers (Buckland et al. 2001; Thomas et al. 2010); therefore, it appears better suited to my dataset given the strong evidence for heaping. Of the models tested for this data filter, the half-normal key model with cosine adjustments had the lowest AIC score (Table 2.1).

Table 2.2. Summary of the estimated density (D) and the coefficient of variation (*cv*) for two truncation values (*w*). The data were analyzed ungrouped and with two different groupings: five intervals of equal widths with cutpoints away from favored values, and eight intervals of equal widths. For each data type, an asterisk indicates the model with the smallest AIC score. Models that failed are indicated by "n.a."

| | | Truncation | | | |
|---|---|---|---|---|---|
| | | *w* = 40 m | | *w* = largest distance | |
| Data type | Model (Key + adjustment) | D | cv (%) | D | cv (%) |
| Ungrouped | Half normal + cosine | 1107.9 | 14% | 703.1 | 11% |
| | Half normal + Hermite | n.a. | n.a. | n.a. | n.a. |
| | Uniform + cosine* | 880.8 | 12% | 410.0 | 49,069% |
| | Hazard rate + cosine | 23,071.2 | 19% | 14,654.2 | 2,095,096% |
| Grouped (5 equal) | Half normal + cosine* | 954.7 | 13% | 553.9 | 10% |
| | Half normal + Hermite | 401.1 | 9% | 313.0 | 9% |
| | Uniform + cosine | 889.8 | 15% | 163.7 | 8% |
| | Hazard rate + cosine | 2,329.3 | 75% | 428.0 | 13% |
| Grouped (8 equal) | Half normal + cosine | 1,940.6 | 12% | 850.2 | 11% |
| | Half normal + Hermite | 452.9 | 9% | 315.7 | 8% |
| | Uniform + cosine | 1,372.2 | 11% | 468.9 | 9% |
| | Hazard rate + cosine * | 9,214.3 | 102% | 792.2 | 20% |

The *P* values associated with the $X^2$ goodness-of-fit tests were highly significant across models for this data filter, but three orders of magnitude lower for the half-normal key model with cosine adjustments ($X^2 = 5.14$, d.o.f = 1, $P = 0.023$). The quantile-quantile plot (Fig. 2.4) revealed poor fit near to zero, such that more empirical detections were made than predicted by the model. Heaping was also apparent in the quantile-quantile plot.

Based on the data filter with 5 equal cutpoints away from favored numbers and truncation at 40 meters, and the half-normal model with cosine adjustments, the probability of detection (mean ± SE) is 0.105 ± 0.012 (CI: 0.084, 0.131) within 40 m, the density is 955 ± 122 (CI: 743,1227), and global abundance is 314,100 ± 40,267 (CI: 244,380, 403,700) within the 329 $km^2$ survey area. The detection function and the probability of detection plotted over histograms of the data are given in Fig. 2.3C & D, respectively.

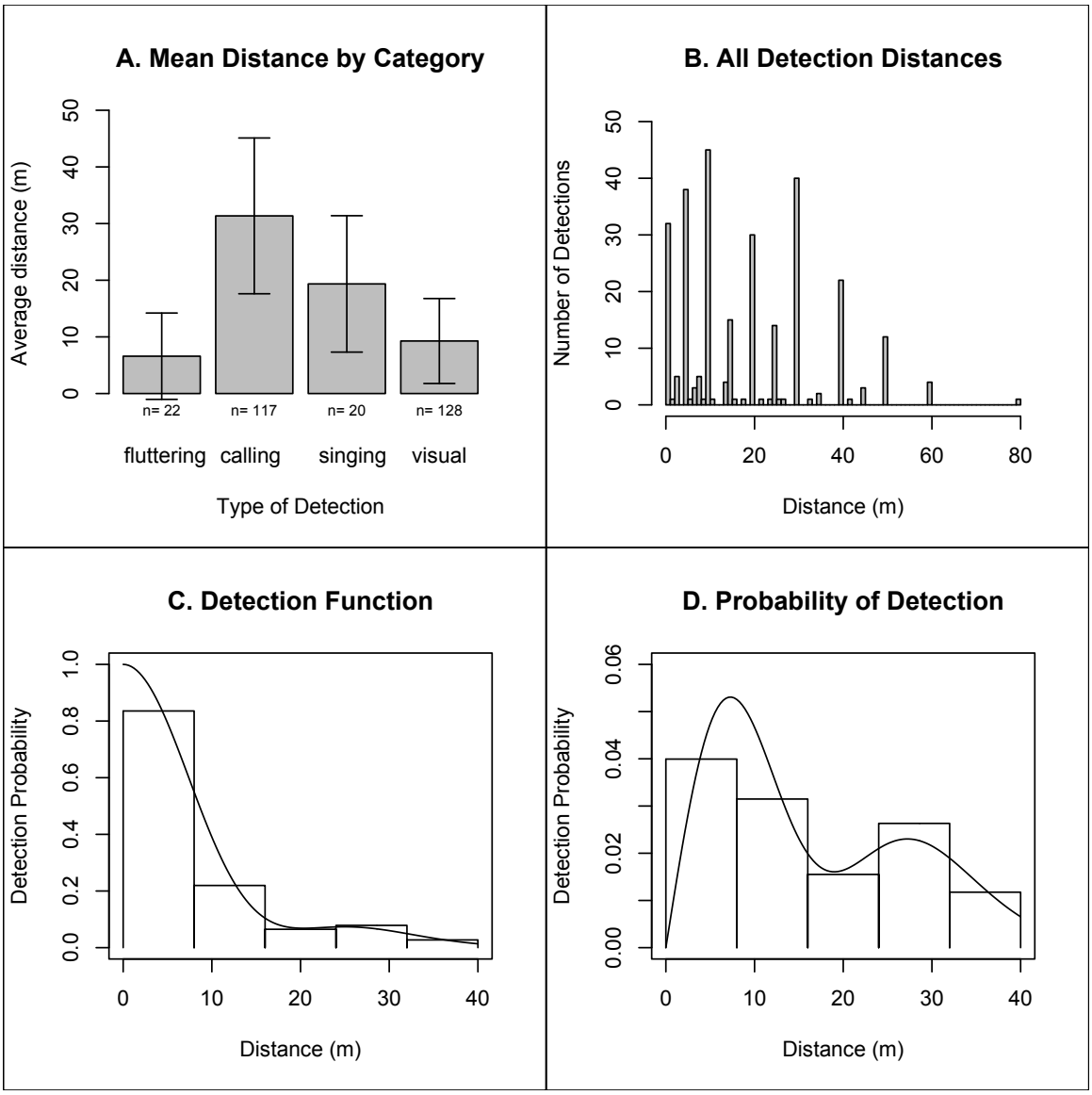



Fig. 2.3 A. Mean distance to the black-billed streamertail in four detection categories: fluttering, calling, singing, and visual. Number (n) of detections in each category is given. B. Histogram of ungrouped distance data. Tendency to round to favored distances (i.e. heaping) is apparent, as is spiking near to zero. Some evidence for over-dispersion to mid-range distances is also apparent. C. Detection function fitted to data using a half normal key model plus cosine adjustments. D. Probability of detection (PDF).

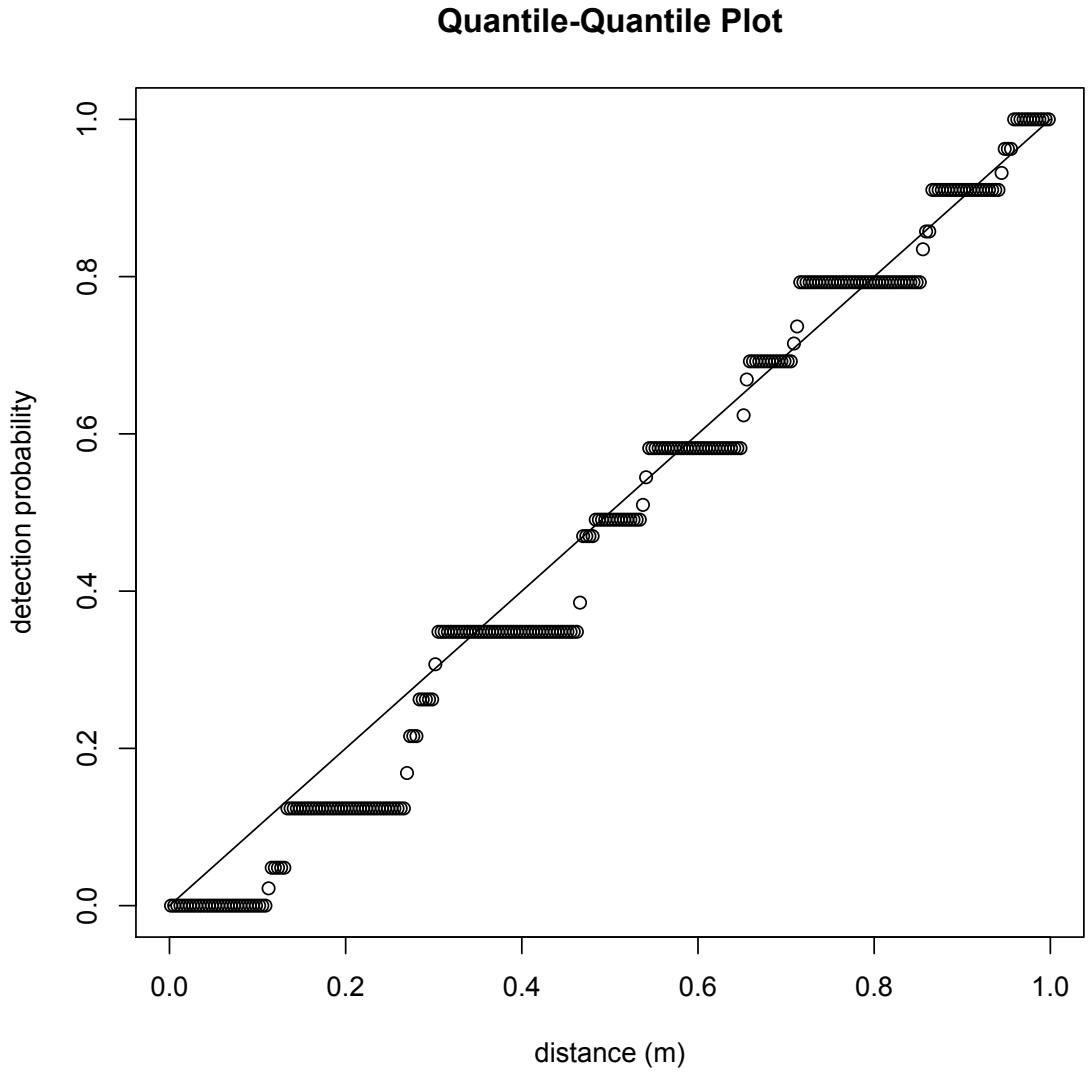



Fig. 2.4. Quantile-quantile plot corresponding to the fit of a half normal detection function model with cosine adjustments. "Stepped" appearance relates to heaping at favored distances, and more detections than expected near g(0) suggests poor model fit.

*Variable Circular Plot*

Following the methods in Reynolds et al. (1980), I summarized detections by 5 m bands around the points. The inflection point occurred at 10 m (Fig. 2.5). Because of the issue with spiking, I conservatively set the effective census radius to 20 m, leaving a total of 187 individuals at 145 points with radius 20 m, or 1,026 individuals/km$^2$ of occupied

habitat. Given an occupancy rate of 32%, the calculation for abundance is 108,017 individuals within the 329 $km^2$ survey region.

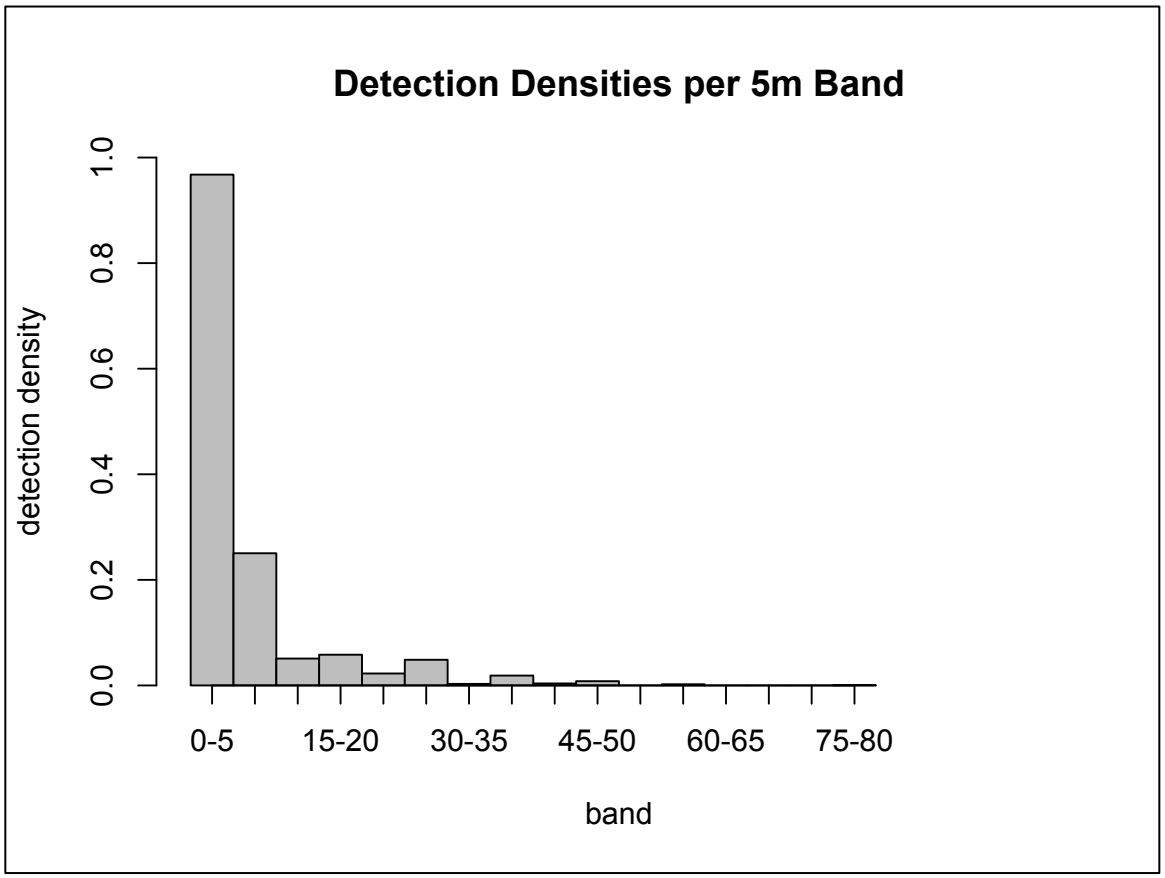



Fig. 2.5. Detection densities for each five-meter band away from points. The inflection point of Reynolds (1980) occurs at ten meters.

**DISCUSSION**

In the current study, the black-billed streamertail density estimated by distance sampling (955 individuals per $km^2$) and the variable circular plot method (1,026 individuals per $km^2$) are roughly two orders of magnitude higher than the 3–6 pair per $km^2$ estimated by Schuchmann (1999). The difference could reflect population growth, seasonal effects, survey method, or any combination of these three factors. Black-billed streamertail occurrence in habitats such as gardens and churchyards suggests that they adapt well to

human modifications of the landscape, and their populations may actually be increasing in low- and mid-elevations where agriculture and urban areas are expansive. The ability of the red-billed streamertail to feed on a wide variety of introduced nectar sources such as *Syzygium jambos* and *Spathodea campanulata* was noted by (Lack 1976). Throughout the survey period, I frequently witnessed the black-billed streamertail feeding on these and other introduced plants. Hence, human activities, in some cases, may serve to increase habitat quality for the black-billed streamertail by increasing the availability of nectar sources.

It is perhaps surprising that any land bird population on Jamaica or elsewhere in the Caribbean could thrive given the frequency and severity of hurricanes to occur there. Documented declines of several Caribbean bird species have occurred as result of hurricanes and major storms (Raffaele 1977; Jeggo and Taynton 1980; Smith and Temple 1982; but see Varty 1991). However, Hurricane Gilbert, which devastated parts of Jamaica in 1988, did not cause drastic population declines for several species of land birds surveyed, at least in the short- to immediate-term (Varty 1991). Wunderle et al. (1992) compared population densities of land birds on Jamaica before and after the hurricane, and showed that nectivorous and frugivorous bird populations (including the red-billed streamertail) had sharper declines in local density after Hurricane Gilbert (1988) than did insectivorous bird populations (Wunderle et al. 1992). However, the general pattern among localities was idiosyncratic: mean number of birds declined in two montane habitats, but increased in two lowland sites, and stayed the same in the remaining five lowland sites (Wunderle et al. 1992). Movement, rather than mortality, may have caused the majority of the change in local abundance. Black-billed streamertail

populations might be resilient to the adverse effects of hurricanes if they are able to move away from areas of heavy defoliation to avoid starvation. Given their high vagility and generalist foraging strategy (feeding on a wide variety of plants), this is likely to be the case. Post hurricane observations of the black-billed streamertail in neighboring St. Andrew Parish (or approximately 20 - 25 km west of the known distribution) suggest they were either blown or flew out of their typical range in order to find nectar (Gosse Bird Club 1988, 1989). However, more studies are needed to better understand the impacts of human activity and natural disturbances on the black-billed streamertail densities in different habitats.

The reliability of density and abundance estimates obtained with conventional distance sampling methods depends heavily on whether or not the data meet four key assumptions. First, distance modeling assumes that birds on the point are certain to be detected. Failure to detect all birds at zero distance from the point will result in a negative bias in the estimate of density and abundance (Buckland et al. 2001, 2008). The montane wet forests of the John Crow and Blue Mountains, like all tropical forests, have a complex vertical structure and low light levels. Both sexes can be fairly camouflaged against a green background in low light conditions. The inconspicuously plumed females can be especially difficult to detect. Therefore, I cannot rule out the possibility that some birds at zero distance from the point were missed.

Second, objects must be detected at their initial location. The mathematical theory underlying distance sampling assumes that random movement of objects does not occur. In reality, all birds move, which creates a bias because the probability of detection is a non-increasing function of distance from the point. Objects moving at random are more

likely to be detected when moving closer to the point, leading to an underestimate of distance and a positive bias in density estimation. In point counts, where the observer is stationary, overestimation due to random movement is especially problematic (Buckland et al. 2001). Some published studies have shown that density estimates can be inflated by as much as a factor of ten due to these random effects (Bibby et al. 2000). A snapshot approach, in which distance to detected birds are measured in a snapshot moment (Buckland et al. 2001, Buckland 2006, Buckland et al. 2008), may help mitigate the positive bias caused by random and responsive movement during the counting period. This approach may prove effective for future surveys of the black-billed streamertail.

Movement of objects can also be non-random in response to the observer. Response to the observer may take the form of movement toward or away from the observer. Responsive movement away from the observer would lead to underestimation of density, whereas movement toward to the observer would lead to overestimation of density (Bibby et al. 2000, Buckland 2006, Thomas et al. 2010). The black-billed streamertail did not tend to exhibit highly evasive behavior, such as flushing. For example, in most cases, individuals that were singing continued to sing, foraging birds continued to feed, and dominant individuals appeared to be completely consumed with chasing away competitors, thus paying the observers scant attention. For the current study, bias arising from evasive movement might be small, though bias arising from random movement may be large.

The third assumption of distance modeling is that distances to detected individuals are precisely measured (Buckland et al. 2001, Buckland 2006, Buckland et al. 2008, Thomas et al. 2010). If distances were underestimated by 10%, then densities

would be overestimated by 100 * ($1/0.9^2$ - 1) = 23%; if distances were overestimated by 10%, then densities would be underestimated by 100 * (1 - $1/1.1^2$) = 17% (Buckland et al. 2001). In the current study, the use of a laser range finder increased the precision of distance estimates to visually detected objects. However, distance estimates for aural detections can only be approximated. Breeding bird surveys conducted in forest habitats can have up to 90% aural detections (Reynolds et al. 1980, Bibby et al. 2000). Somewhat surprisingly, only 55% percent of black-billed streamertail detections were aural, thereby limiting impact of biased estimation relative to surveys for which there are higher percentages of aural detections. However, biased estimation could still be an issue for this survey. Examining the histogram of detection distances (Fig. 2.3 B), there appears to be some signal for over-dispersion, and the probability density function (Fig. 2.3 D) tries to fit a smaller peak at around 30 m. Given that hummingbirds do not typically flush, this pattern could be the result of imprecise distance measurements, or an artifact caused by spiking.

Finally, the sampled plots must be representative of the entire region. Generally speaking, surveys conducted along roads and paths constitute a poor survey design and may bias survey data (Marques et al. 2007; but see Rivera-Milán et al. 2015 for an exception). I surveyed primarily along roads and community footpaths rather than using a randomized survey design or grid. If Otaheite Apple (*Syzgium malaccense*), hibiscus (*Hibiscus rosa-sinensis*), "Grow stick" (*Gliricidia sepium*), and other major nectar sources are planted preferentially along roads, then streamertail density may be higher along roads than non-roads and would produce an upward bias in population estimate. Apart from planted nectar sources, the openness of human-modified forests, housing, or

agricultural areas might represent a structural advantage for this species. Schuchmann (1999) notes that densities of streamertail hummingbirds are lower in extremely thick vegetation. In reality, there is no way to avoid this potential source of bias because of the logistical constraints and the difficult terrain.

Mischaracterization of the larger region can also occur if census points are clustered too tightly together, which can lead to counting the same individuals across multiple census points. While counting the same bird across multiple survey points does not violate the assumptions of distance sampling *per se*, tight clustering can limit the survey's ability to detect variation in density across the larger region. By spacing the survey points at least 200 m apart, I minimized this source of bias. Because I was driving between census points for most of the survey (approximately 90%), instances of double counting are likely rare. Overall, the survey points were well spaced throughout the larger region (excluding inaccessible areas; Fig. 2.2).

The 'spike' in the black-billed streamertail data can indicate a failure of model assumptions or a real biological feature (Thomas et al. 2010). Only adult males produce the fluttering noise in flight. Therefore, some spiking may have occurred because there are more ways to aurally detect adult males at small distances from the point. Spiking can also arise if animals move toward the observer, though, as discussed above, the black-billed streamertail does not tend to display responsive behavior. Finally, rounding errors at small distances can also cause a spike in the data.

The fact that the density estimated using the variable circular plot method (1,026 individuals per $km^2$) is similar to the density estimated using these distance methods is problematic because the density for the variable circular plot method is based on

occupied points only, while the density for the distance methods estimate is based on both occupied and unoccupied points. This fact, together with the issues listed below for model fit (spiking, rounding, etc.), suggests that the distance sampling analysis may have a positive bias, and the results should be interpreted with caution. The smaller estimate of total abundance using the variable circular plot, which is adjusted for occupancy rate (108,017 individuals), may be more realistic.

**CONCLUSIONS**

The global estimate for black-billed streamertail abundance under both the variable circular plot method (108,017 individuals) and distance methods (mean ± SE, 314,100 ± 35,427 (CI: 251,720 - 391,930)) support the current IUCN status of Least Concern v 3.1. The more conservative global abundance estimate of 108,017 individuals (variable circular plot method) may be more reliable due to the potential overestimation of abundance in the distance sampling analysis. Ongoing monitoring of black-billed streamertail is recommended to better understand population trends, habitat preferences, and the impact of human and natural disturbance events.

## REFERENCES


Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pp. 267–281 *in* B. Petrov and F. Csaki, eds.

Bibby, C. J., N. D. Burgess, D. A. Hill, and S. Mustoe. 2000. Bird census techniques. Second. Academic Press, San Diego.

Bird Life International 2012. *Trochilus scitulus*. The IUCN red list of threatened species. Version 2014.1. Downloaded on 15 July 2014.

Brewster, W., and O. Bangs. 1901. On an overlooked species of *Aithurus*. Proceedings of the New England Zoölogical Club 2:47–50.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling. Estimating abundance of biological populations. Oxford University Press, Oxford.

Buckland, S. T., S. J. Marsden, and R. E. Green. 2008. Estimating bird abundance: making methods work. Bird Conservation International 18:S91–S108.

Clark, C. J. 2008. Fluttering wing feathers produce the flight sounds of male streamertail hummingbirds. Biology Letters 4:341–344.

Coyne, J. A., and T. D. Price. 2000. Little evidence for sympatric speciation in island birds. Evolution 54:2166–2171.

Fewster, R. M., S. T. Buckland, K. P. Burnham, D. L. Borchers, P. E. Jupp, J. L. Laake, and L. Thomas. 2009. Estimating the encounter rate variance in distance sampling. Biometrics 65:225–236.

Gill, F. B., F. J. Stokes, and C. Stokes. 1973. Contact zones and hybridization in the Jamaican hummingbird, *Trochilus polytmus* (L.). The Condor 75:170–176.

Gosse, P. H. 1847. The birds of Jamaica. Van Vorst, London.

Gosse Bird Club 1988. Observations on the effects of hurricane Gilbert on bird life in Jamaica. Gosse Bird Club Broadsheet 2–7.

Gosse Bird Club 1989. Post-hurricane observations. Gosse Bird Club Broadsheet 8–9.

Graves, G. R. 2015. A primer on the hybrid zone of Jamaican streamertail hummingbirds (Trochilidae: *Trochilus*). Proceedings of the Biological Society of Washington 128:111–124.

Graves, G. R. 2009a. Ontogeny of bill color in male streamertail hummingbirds (*Trochilus*). Journal of Caribbean Ornithology 44–47.

Graves, G. R. 2009b. Skeletal correlates of body weight in the black-billed streamertail (*Trochilus scitulus*) of Jamaica. Caribbean Journal of Science 45:69–72.

Jeggo, D. F., and K. M. Taynton. 1980. The effects of hurricane Allen on the status of the St. Lucia parrot, *Amazona versicolor*. Dodo 17:11–18.

Lack, D. 1976. Island biology illustrated by the land birds of Jamaica. XVI. University of California Press, Berkeley and Los Angeles.

Lance, S. L., C. Hagen, T. C. Glenn, R. T. Brumfield, K. F. Stryjewski, and G. R. Graves. 2009. Fifteen polymorphic microsatellite loci from Jamaican streamertail hummingbirds (*Trochilus*). Conservation Genetics 10:1195–1198.

Linnaeus, C. 1758. Systema naturæ. 10th edition. Holmiae : Impensis direct. Laurentii Salvii., 823 pp.

Marques, T. A., L. Thomas, S. G. Fancy, S. T. Buckland, and C. M. Handel. 2007. Improving estimates of bird density using multiple-covariate distance sampling. The Auk 124:1229–1243.

McCormack, J. E., J. M. Maley, S. M. Hird, E. P. Derryberry, G. R. Graves, and R. T. Brumfield. 2012. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Molecular Phylogenetics and Evolution 62:397–406.

Raffaele, H. A. 1977. Comments on the extinction of *Loxigilla portoricensis grandis* in St. Kitts, Lesser Antilles. Condor 79:389–390.

Reynolds, R. T., J. M. Scott, and R. A. Nussbaum. 1980. A variable circular-plot method for estimating bird numbers. The Condor 309–313.

Rivera-Milán, F. F., P. Bertuol, F. Simal, and B. L. Rusk. 2015. Distance sampling survey and abundance estimation of the critically endangered Grenada dove (*Leptotila wellsi*). The Condor 117:87–93.

Schuchmann, K. L. 1999. Handbook of the birds of the world. Lynx Edicions, Barcelona.

Schuchmann, K. L. 1980. Die Jamaika kolibris *Trochilus polytmus* und *Trochilus scitulus*. Biotropic-Verlag, Frankfurt am Main.

Schuchmann, K. L. 1977. Notes on the display of the streamertailed hummingbird *Trochilus polytmus* (Linn.). Gosse Bird Club Broadsheet 28:11–13.

Smith, T. B., and S. A. Temple. 1982. Grenada hook-billed kites: recent status and life history notes. Condor 84:131.

Thomas, L., S. T. Buckland, E. A. Rexstad, J. L. Laake, S. Strindberg, S. L. Hedley, J. R. B. Bishop, T. A. Marques, and K. P. Burnham. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. Journal of Applied Ecology 47:5–14.

Varty, N. 1991. The status and conservation of Jamaica's threatened and endemic forest avifauna and their habitats following hurricane Gilbert. Bird Conservation International 1:135–151.

Wunderle, J., Joseph M., D. J. Lodge, and R. B. Waide. 1992. Short-term effects of hurricane Gilbert on terrestrial bird populations on Jamaica. The Auk 109:148–166.

## CHAPTER 3.

## MORPHOLOGICAL AND GENETIC CHARACTERIZATION OF THE HYBRID ZONE BETWEEN JAMAICAN-ENDEMIC STREAMERTAIL HUMMINGBIRDS (*TROCHILUS POLYTMUS* AND *T. SCITULUS*)

### INTRODUCTION

Some of the most dramatic examples of avian radiations occur in ocean archipelagos where oceanic barriers to dispersal and small population sizes facilitate divergence between populations on different islands. In contrast, few examples of avian speciation occur *in situ* due to their limited geographic area. A survey of avian sister species inhabiting oceanic islands and small archipelagos failed to detect a single case that could not be explained by secondary invasions or inter-island allopatric speciation (Coyne and Price 2000). For example, the spectacular radiations of Hawaiian honeycreepers and Darwin's finches are thought to have arisen through multiple instances of allopatric speciation on different islands followed by secondary colonization.

The distinctive hummingbird genus *Trochilus*, which is represented exclusively by two sexually dichromatic taxa endemic to the oceanic island of Jamaica, has long represented a possible exception to the rule that *in situ* speciation on small oceanic islands does not occur in birds (Coyne and Price 2000). At approximately 11,400 $km^2$, Jamaica is the smallest oceanic island in the Western Hemisphere. While they possess a long shared history on Jamaica, a phylogenetic analysis of all hummingbirds based on the mitochondrion and several nuclear loci revealed only shallow divergence between them (McGuire et al. 2014). Yet they are distinguished by male bill color (Brewster and Bangs 1901), a secondary sexual ornament, which is coral red in the red-billed streamertail (*Trochilus polytmus*), and jet black in the black-billed streamertail (*T. scitulus*). The bill

of *T. polytmus* is also wider than in *T. scitulus* (Brewster and Bangs 1901; Graves 2015). The redness of the bill does not appear to be a pigment because it fades rapidly post mortem. More likely, the red color is due to the presence of heavily vascularized tissue that is visible on the portions of the bill lacking melanin. More subtle differences in bill length, tail length, and degree of tail forking have been documented (Graves 2015), as have some differences in song and courtship displays (Schuchmann 1978, 1980). Both taxa have what appears to be a year-round lek mating system (Gosse 1847), with males performing displays that emphasize their bill color, the dramatically elongated second to outermost tail feathers, and horned crown feathers (Gill et al. 1973; Schuchmann 1978, 1980).

Hybrids were first documented nearly 70 years ago, and a narrow hybrid zone was discovered where their parapatric ranges meet along the northern side of the island (reviewed in Graves 2015). The geographic extent of the hybrid zone was mapped by Gill et al. (1973) nearly 50 years ago. The hybrid zone appears to be centered on the Rio Grande Valley (Gill et al. 1973, Graves 2015). Since then, survey and collection efforts indicate that hybrid zone has not moved or widened significantly during this half century (MacColl and Lewis 2000, Graves 2009a, 2009b, 2015, Judy 2018). Along the southern end of the island, the contact zone is found along the Morant River, but no hybrid zone occurs there (Gill et al. 1973).

Hybrid zones, where genetically divergent forms meet and mate (Harrison 1993), are natural laboratories to study evolution (Hewitt 1988). Careful examination of patterns of morphological and genetic variation across a hybrid zone can shed light on the selective forces maintaining it (Barton and Hewitt 1985; Hewitt 1988). Clines can be

dispersal-independent for sedentary organisms (Moore 1977), or dispersal-dependent for vagile organisms where the homogenizing effect of dispersal is balanced by spatial heterogeneity (Barton and Hewitt 1985). Dispersal-dependent clines have three categories (Barton and Hewitt 1985; Hewitt 1988): 1) neutral clines where initially steep gradients decay with time in the absence of selection against hybridization, 2) waves of advance of an advantageous allele from one species into another (Ferris et al. 1983), and 3) equilibrium zones termed 'tension zones' (Key 1968) that are maintained by a balance between selection against hybrids and the dispersal of parental types into the zone (Haldane 1948; Barton 1979). Care must be taken to distinguish among these models to make correct inference about selection, because they can produce the same clinal patterns.

Cline theory provides a conceptual and statistical framework for comparing geographic cline shapes and positions for individual traits in order to distinguish among cline models and make inferences regarding selection (Barton 1979; Barton and Hewitt 1989; Harrison 1993). For instance, traits that do not introgress freely are expected to have narrow clines relative to traits that do (Barton 1979; Szymura and Barton 1986). Cline centers should coincide with the approximate geographic location for the majority of traits examined. However, the cline center for an individual trait can be shifted as a result of genetic drift, positive selection, or artificially, through allele dominance relationships or gene interactions (Butlin 1994, Brumfield et al. 2001, Baldassarre et al. 2014). For independently assorting loci, varying cline widths reflect the relative strength of selection acting against the locus in hybrids (Butlin et al. 1991).

Here, I used a large series of *Trochilus* specimens (n = 201) to characterize the morphological and genetic structure of the hybrid zone. My specific goals are (i) to characterize variation within and between species in two bill traits (color and width) and body size, (ii) determine whether the two species form discrete genomic clusters based on genotypes of six microsatellite loci and 6,451 single nucleotide polymorphisms (SNPs), and (iii) generate cline models for morphological and genetic traits that show strong differences between species. I expect steeper clines to be associated with stronger divergent selection, and wider clines to be associated with weaker selection.

## METHODS

### Sample collection

All voucher specimens and tissues used in this study (Table S3.1) were collected under permits issued by Jamaica's National Environment and Planning Agency (NEPA), Kingston. From November 2003 to March 2006, GRG netted 171 *Trochilus* individuals along a west-to-east transect spanning the transition from pure red-billed *T. polytmus* to pure black-billed *T. scitulus* (Fig. 3.1).

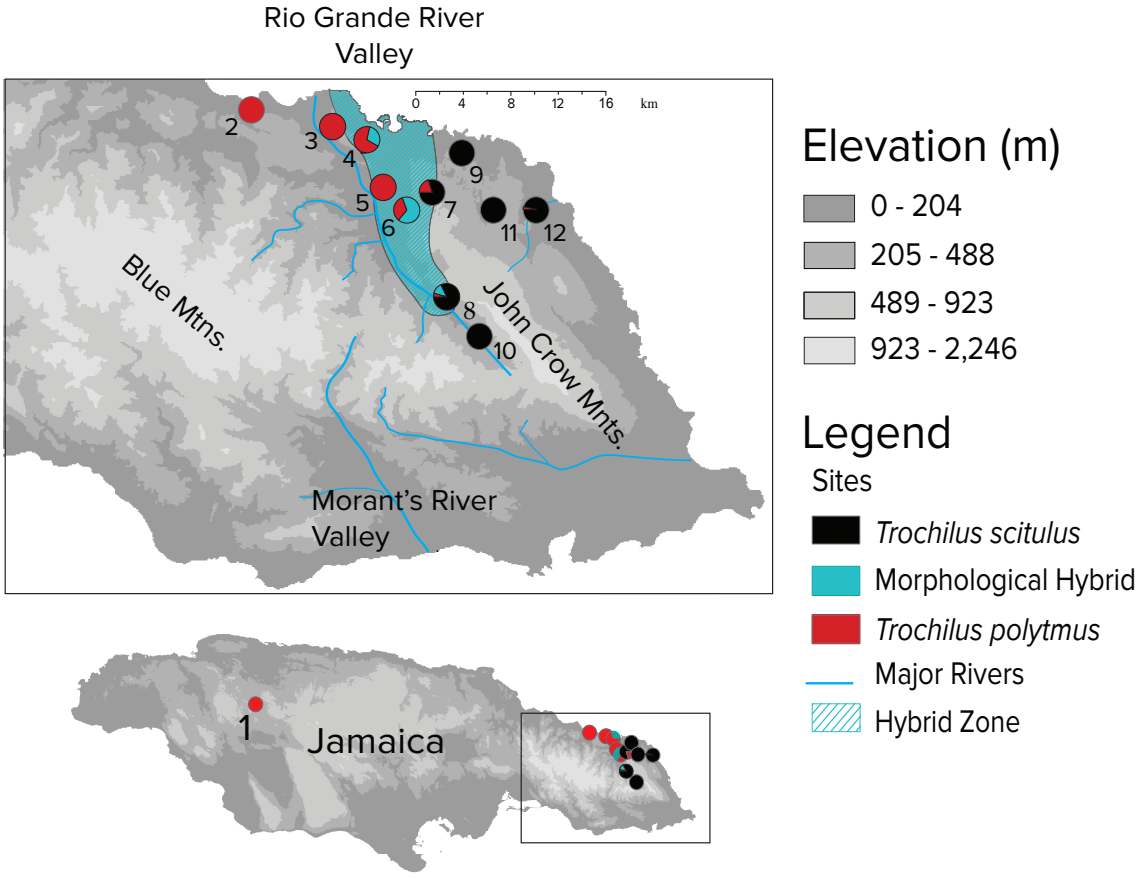



Fig. 3.1. Sites 1-12 across Jamaica, West Indies. The hybrid zone indicated by the teal shaded region overlaps with the Rio Grande Valley in Eastern Jamaica. See inset map for Site 1.

Freshly euthanized birds were weighed using a digital scale, and two separate digital photographs of the bill in dorsal aspect were taken. Additionally, details were recorded for each specimen regarding the time and place of collection, sex, age, plumage and soft part coloration, breeding condition, and type of tissues preserved. In December 2014, I netted an additional 26 *Trochilus* individuals along the same transect using the same protocol. Specimens are deposited in the research collection of the U. S. National Museum of Natural History (NMNH) Smithsonian Institution, in Washington D.C., the

Louisiana State University Museum of Natural Science (LSUMNS) in Baton Rouge, LA, or the Institute of Jamaica (IOJ) in Kingston, Jamaica.

**Phenotypic measurements**

*Bill color*

I measured bill color using photographs of living or freshly euthanized individuals (see methods in Graves 2009a, 2015). Briefly, individuals were placed in dorsal aspect next to a millimeter scale, and were visually assessed to determine the extent of black on the dorsal surface of the maxilla. Various hybrid indices have been developed to describe bill color in *Trochilus* (Brewster and Bangs 1901; Gill et al. 1973; Graves 2009a, 2015). I use a five-category index devised by Graves (2015), which yields repeatable classification from photographs. For males, category 1 corresponds to entirely black bills, and category 5 corresponds to bills that are bright red and narrowly tipped in black. The hybrid index for females is similar, though female *T. polytmus* have duller bills than males. For both sexes, birds scored as 2-3 were defined as hybrids (Fig. S3.1).

I tested for interspecific differences in bill color using reference sample sites (Table 3.1). I excluded from these tests two *T. polytmus* individuals (MLB 369 and MLB 372) that were netted at Site 12 (discussed below). I used a nonparametric Wilcoxon rank sum test to compare bill color, a quantitative discreet character defined by the hybrid index. Because the hybrid indices used for males and females were derived independently and cannot be directly compared (i.e. a female "4" is inherently different than a male "4"), I did not make intra-specific comparisons, but focused instead on interspecific comparisons between male *T. polytmus* and *T. scitulus*, and between female *T. polytmus* and *T. scitulus*.

| Site No. | Locality Name | Distance from Site 1 (km) | North | West | *n* | Type of locality | Bill width (mean) | Bill width (S.D.) | Hybrid index (mean) | Hybrid index (S.D.) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Windsor | 0.0 | 18.34 | -77.67 | 16 | ref. *T. polytmus* | 4.71 | 0.39 | 4.18 | 0.39 |
| 2 | Somerset | 119.2 | 18.20 | -76.55 | 5 | ref. *T. polytmus* | 4.38 | 0.29 | 4.88 | 0.35 |
| 3 | Burlington | 125.5 | 18.18 | -76.50 | 10 | *T. polytmus* | 4.43 | 0.12 | 4.50 | 0.52 |
| 4 | Springbank | 128.2 | 18.18 | -76.47 | 10 | hybrid zone | 4.33 | 0.08 | 4.71 | 0.49 |
| 5 | Fellowship | 130.1 | 18.14 | -76.46 | 5 | hybrid zone | 4.32 | 0.16 | 4.44 | 0.73 |
| 6 | Tom's Hope | 132.1 | 18.13 | -76.44 | 10 | hybrid zone | 4.10 | 0.22 | 2.53 | 1.37 |
| 7 | Trowel Hill | 133.8 | 18.13 | -76.43 | 10 | hybrid zone | 4.01 | 0.25 | 1.41 | 1.18 |
| 8 | Comfort Castle | 136.7 | 18.05 | -76.41 | 28 | hybrid zone | 3.99 | 0.19 | 1.61 | 1.17 |
| 9 | Nonsuch | 135.2 | 18.16 | -76.41 | 14 | *T. scitulus* | 3.93 | 0.10 | 1.00 | 0.00 |
| 10 | Millbank | 139.9 | 18.03 | -76.39 | 7 | ref. *T. scitulus* | 3.79 | 0.18 | 1.00 | 0.00 |
| 11 | Cambridge Barracks | 138.2 | 18.12 | -76.38 | 3 | ref. *T. scitulus* | 3.88 | 0.18 | 1.00 | 0.00 |
| 12 | Ecclesdown | 143.2 | 18.10 | -76.34 | 12 | ref. *T. scitulus* | 3.86 | 0.36 | 1.33 | 0.90 |

Table 3.1. Site information including site number, locality name, the linear distance from Site 1 in kilometers, the geographic coordinates in decimal degrees, the number of sampled individuals (*n*), the type of locality as defined for analysis (ref. = reference, or hybrid zone), mean and standard deviation (S.D.) for bill width in millimeters, and the mean and standard deviation for hybrid index.
Fig. 3.1. Sites 1-12 across Jamaica, West Indies. The hybrid zone indicated by the teal shaded region overlaps with the Rio Grande Valley in Eastern Jamaica. See inset map for Site 1.

*Bill width*

I assessed bill width from digital photographs (see methods in Graves 2009a, 2015). Briefly, bill width was measured to the nearest 0.01 mm at the anterior extension of feathers on the dorsal surface of the mandible from enlarged images (30X) on all specimens collected by GRG. I measured all specimens collected in 2014 following methods in Graves (2009a, 2015) but used digital landmarks in the program TPSDig v. 2.05 (Rohlf, 2016). The use of digital photographs on living or freshly euthanized birds circumvents distortion issues that arise during specimen preparation and the process of drying itself (Graves 2015). Bill distortion can be especially pronounced in hummingbirds because the ventral bar of the maxilla is fragile and can flex when tied shut (Zusi 2013; Graves 2015).

I made two independent measurements of bill width, one from each of the two photographs for each bird. I report the average of the two measurements for bill width for each adult individual (i.e. those lacking obvious bill striations) for which bill width measurements were available (Table S3.1). I tested for intra- and interspecific differences in bill width using reference sites (Table 3.1). I excluded from these tests the two *T. polytmus* individuals that were netted within the *T. scitulus* reference sites. Individuals were divided into four groups: *T. polytmus* males, *T. polytmus* females, *T. scitulus* males, and *T. scitulus* females. To determine whether bill width trait data met the assumptions for a one-way ANOVA, I tested for normality using a Shapiro-Wilk test and homogeneity of variance using a modified Levene Test based on the absolute deviations from the trimmed mean with a correction factor. No significant departures from normality or deviations from homogeneity of variance were detected.

*Body size*

I examined size variation of three skeletal elements: sternum length, keel length, and keel depth, to explore potential body size differences between species, and sexual size dimorphism within species. Direct measures of body mass on live or freshly euthanized birds can be heavily biased by gut contents (Graves 2009b, 2015). This bias can be significant in small birds like hummingbirds, making it difficult to assess minute differences between similarly sized species. GRG measured the specimens collected in 2003 – 2006, and I measured the specimens collected in 2014. Additionally, I measured five specimens previously measured by GRG to assess the inter-observer measurement error. I used all specimens for which skeletal elements were available, excluding hybrids and individuals with noted bill striations, an indicator of juvenile status (Ortiz-Crespo 1972). Prior to performing an ANOVA, I tested for normality using a Shapiro-Wilk test. I tested for homogeneity of variance using a modified Levene Test based on the absolute deviations from the median with no correction factor.

**Molecular methods**

*Microsatellites*

Fragment length differences were determined in nuclear microsatellite loci for 172 individuals across 12 sample localities using nine *Trochilus*-specific autosomal microsatellite markers (Lance et al. 2009). One primer from each pair was modified on the 5' end with an engineered sequence (M13) to enable use of a third primer in the PCR (identical to the M13 tag) that was fluorescently labeled for detection. PCR amplification of fragments were multiplexed in a 10 μL reaction volume (0.01 μM forward primer, 0.22 μM reverse primer, 0.02 dye-labeled universal primer, 10 mg/mL BSA, 5 units Qiagen

Multiplex Mix, and 20 ng DNA template) using a Peltier Thermal Cycler 200 DNA Engine Cycler with the following cycling parameters: initial denaturation at 95° for 15 min, 24 cycles of 95° for 30 s, 59° for 90 s, and 72° for 60 s, followed by 19 cycles of 94° for 30 s, 72° for 90 s, and 72° for 60 s, and a final extension at 60° for 30 min. PCR products were resolved on an Applied Biosystems (AB) 3100 capillary sequencer and scored by size using GeneMapper (AB) software version 4.01.

*Genotyping-by-sequencing*

DNA extracts for 160 *Trochilus* individuals were sent to Cornell Institute of Genomic Diversity (IGD) to generate a genome-wide SNP dataset following the genotyping-by-sequencing method (Elshire et al. 2011). Briefly, DNA from each individual was digested with PstI (CTGCAG). Sample-specific adapters and common adapters were ligated to the resulting fragments. Fragments were then pooled and cleaned using a QIAquick PCR purification kit (Qiagen, Valencia, CA, USA) and amplified using an 18-cycle PCR. The amplified libraries were purified using QIAuquick columns and quantified using a PicoGreen assay (Molecular Probes, Carlsbad, CA, USA). Final library products were sequenced as single-end reads to 100-base pairs on an Illumina HiSeq 2000 lane (Illumina, San Diego, CA, USA).

SNP calling was performed using UNEAK, a reference-free SNP calling pipeline specific to GBS data that is an extension of the Java program TASSEL (Bradbury et al. 2007). We chose the UNEAK pipeline because it calls nearly as many SNPs as reference-based pipelines with high accuracy (Torkamaneh et al. 2016). UNEAK performs an initial filter to remove reads that lack either a barcode or a cut site, or has ‘N’s present in the first 64 bases of the sequence after the barcode. Reads that pass this initial filter are

trimmed to 64 bp (including the cut site remnant but removing the barcode). If a read contains a full cut site or part of an adapter, the read is trimmed appropriately and padded to 64 bases with polyA. Identical reads are clustered into tags, and counts of these tags per individual are stored. Then, all unique tags are merged and their counts in the whole sample of individuals are stored. Pairwise alignments of tags with a 1-bp mismatch were considered as candidate SNPs. To remove rare or singleton tags that likely result from sequencing error, the UNEAK pipeline then filters SNPs for which the minor allele was represented in fewer than five reads or has an overall frequency less than 5% in the total sample. Finally, the network filters for only reciprocal tag pairs using an error tolerance rate of 0.03 (recommended for Illumina data).

Running the initial UNEAK pipeline with the above settings yielded 97,432 SNPs. I filtered out SNPs with greater than 20% missing individuals per locus, resulting in 32,949 SNPs. Following methods in White et al. (2013), I used a custom perl script (*PairDuplicates.pl*) to filter loci that are likely reverse complements of other loci in the dataset. Finally, we used a second custom perl script (*UNEAK_filter1.pl*) to remove paralogous loci, which we defined as loci having > 0.75 mean heterozygosity. The final call set included 6,451 SNPs.

Two individuals (GRG 4137, GRG 4139) had ambiguous calls ('N's) for the majority of SNPs (> 90%) and were removed from the dataset. I pruned an additional 14 individuals that were outliers in multidimensional scaling and PCA plots that potentially were biased by experimental (sequencing) errors: GRG 3949, GRG 3951, GRG 3976, GRG 3977, GRG 3986, GRG 3989, GRG 4009, GRG 4107, GRG 4115, GRG 4137,

GRG 4145, GRG 4147, GRG 4151, GRG 4159. The final dataset for all downstream analyses contained 6,451 high-quality SNPs from 145 individuals (Table S3.1).

**Analysis of Molecular Data**

*Microsatellites*

I tested all *Trochilus*-specific microsatellite loci for Hardy-Weinberg and linkage equilibrium using randomization procedures implemented in the program FSTAT 2.9.3.2 (Goudet 1995). To test for Hardy-Weinberg equilibrium, alleles were randomized within samples. Tables were classified using the $F_{IS}$ statistic and were based on 1,000 randomizations. To test for linkage equilibrium, genotypes were permuted 360,000 times among samples. I also examined null allele frequencies using an iterative algorithm to compare observed and expected genotype frequencies, which I implemented in CERVUS 3.0 (Kalinowski et al. 2007). In the absence of a null allele, estimated frequencies are expected to be zero or slightly negative.

I quantified gene diversity, an unbiased estimator of heterozygosity (Nei 1973), for each locus-site combination. Allelic richness was also calculated by rarefaction analysis to account for uneven sample sizes across sample localities (Mousadik and Petit 1996; Petit et al. 1998). I examined differentiation among sample localities, as defined in Table S3.1, using a variant of $F_{ST}$, $\Theta$ (Weir and Cockerham 1984). Significant deviations from Hardy-Weinberg and linkage equilibria were evaluated via log-likelihood G tests and randomization procedures, and Bonferroni corrections were applied in cases of simultaneous multiple comparisons of the data (Goudet et al. 1996). I conducted calculations and testing (randomization procedures) using the program FSTAT vs. 2.9.3.2 (Goudet 1995). I also performed an analysis of molecular variance to independently

estimate the among- and within-sample locality components of genetic variance implemented in the program GENALEX 6 (Peakall and Smouse 2006).

I used STRUCTURE v. 2.3.4 (Pritchard et al. 2000), a Bayesian approach to modeling the number of genetic clusters (K), or populations, in the microsatellite dataset. I performed five runs at each K (K = 1 – 4) using a burn-in time of 100,000 followed by 1,000,000 MCMC iterations. Runs were performed under an admixture model with the alleles correlated frequencies option. These options are appropriate for populations that share polymorphisms due to migration or shared ancestry (Falush et al. 2003). Because the underlying model of structure is stochastic, individuals can be assigned to different populations among replicate runs (Jakobsson and Rosenberg 2007). To handle this "label switching" issue, I combined the five runs in CLUMPP using the "FullSearch" option (Jakobsson and Rosenberg 2007), and visualized the results using DISTRUCT (Rosenberg 2004). To determine which K is most probable for my dataset, I examined the posterior probabilities for K = 1 – 4.

*Genotyping-by-sequencing*

I used model-based and multivariate clustering approaches to estimate proportions of ancestry and visualize global patterns of population structure in the GBS dataset. I performed model-based analyses using FASTSTRUCTURE (Raj et al. 2014), which is conceptually similar to the well-known program STRUCTURE (Pritchard et al. 2000), but uses a variational Bayesian framework for posterior inference that makes it run faster with little cost of accuracy (Raj et al. 2014). This algorithm is particularly useful for large SNP datasets. I used the default simple prior and set the number of true populations (K)

to 2. The results were visualized using DISTRUCT2.2 (Vikram Chhatre, 2016: http://www.crypticlineage.net/pages/distruct.html).

I visualized the overall pattern of genetic variability among individuals using a principal component analysis (PCA) using the R programming package, Adegenet (Jombart 2008). In contrast to model-based approaches, multivariate approaches to clustering such as PCAs have the benefit of being able to detect genetic structure in large datasets without making assumptions about the underlying population genetic model (Jombart et al. 2010). To prepare the GBS data for multivariate testing, I scaled and centered the allelic frequencies to mean zero, and I replaced missing data with mean allele frequencies. The PCA was performed using the *dudi.pca* function and the first five axes were retained in the analysis.

To more effectively discriminate between genetic clusters seen in the PCA, I performed a discriminant analysis of principal components (DAPC; Jombart et al. 2010). The DAPC yields synthetic variables of alleles that maximize the among-group variance and minimizes within-group variance (Jombart et al. 2010). The DAPC requires a PCA as a prior step, which transforms the data so that variables are uncorrelated and their numbers are less than that of the analyzed individuals, two key assumptions of a discriminant analysis (Jombart et al. 2010). Here, I tested *a priori* groups defined as *T. polytmus* (hybrid index = 4, 5), *T. scitulus*, (hybrid index = 1), and hybrids (hybrid index = 2, 3).

Retaining too many principal components in the PCA step results in over-fitting, but retaining too few results in insufficient power to discriminate the true biological structures. I used a formal optimization procedure now included in Adegenet (see R

documentation for Adegenet:*xvalDAPC*) to guide my selection of the number of PCs to retain. This "cross-validation" procedure divides the data into a training set (90% of the data) and a validation set (the remaining 10% of the data). The DAPC is conducted on the training set, with variable numbers of PCs retained, and the degree to which the analysis correctly predicts group membership of validation set individuals forms the basis for the optimization. At each level of PC retention, the sampling and DAPC are repeated a user-defined number of times. Cross-validation performed on the *Trochilus* GBS dataset using 30 repetitions showed that the lowest root mean square error was associated with 30 of 144 retained principal components (Fig. S3.6).

I examined the variable contributions (loadings) to the first discriminant function in the DAPC to identify informative SNPs for distinguishing the *Trochilus* species and their hybrids. I define 46 alleles (23 SNP loci) with loadings above 0.001, an arbitrary threshold, to be informative for species discrimination (Fig. S3.7). I calculated allele frequencies differences in parental reference sites for each informative SNP using *Adegenet.*

I used a Bayesian clustering analysis in STRUCTURE (Pritchard et al. 2000) on this informative subset of SNPs under an admixture model with the alleles correlated option. I performed five different runs setting K = 2 and used 100,000 reps as burnin and 1,000,000 MCMC repetitions. Convergence of key parameters (alpha, $F^{ST}$) was assessed by visually inspecting plots of the traces for each parameter. Results were visualized using CLUMPAK (Kopelman et al. 2015), a tool that combines functionality of CLUMPP (Jakobsson and Rosenberg 2007) and DISTRUCT (Rosenberg 2004) into a single step.

## Geographic Cline Analysis

*Phenotypic Traits*

I fit bill color and bill width measurements to five equilibrium trait geographic cline models (sexes separate) using the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm implemented in HZAR (Derryberry et al. 2014). All five models estimate trait variance (center), cline center (distance from site 1, *c*), and cline width (1 / maximum slope, *w*). Models vary with respect to how the exponential tails are fitted: none (no tails fit), both (two tails with independent parameters), mirror (two tails mirrored about the cline center), left (only one fitted tail on the left side of the cline), and right (only one fitted tail on the right side of the cline) (Derryberry et al. 2014). The models were specified in the following manner:

**a.** fixed, none, model I

**b.** fixed, both, model II

**c.** fixed, mirror, model III

**d.** fixed, left, model IV

**e.** fixed, right, model V

I compared models using Akaike information criterion corrected for small sample size (AICc), and considered the model with the lowest AICc as the best-fitting model (Akaike 1973) to allow for comparison among models. HZAR requires that data be collected along a one-dimensional transect (Derryberry et al. 2014). I set the geographic position of each site (1-12) as the centroid (i.e. average) latitudinal and longitudinal coordinates among individuals (sexes pooled) collected at that site. I set the geographic position of each site (1-12) as the centroid (i.e. mean) latitudinal and longitudinal coordinates among

individuals (sexes pooled). I used the westernmost sample locality (Site 1) as the anchor for the analyses, and the geographic positions of the other sample localities were transformed into a list of Euclidean distances from the western anchor using the HZAR functions “hzar.map.latLongSites” and “hzar.map.distanceFromSite.” Placing the anchor further south or north of the centroid position did not have a large effect on estimated cline parameters (Fig. S3.8 A-C, Table S3.2).

I fixed trait mean and variance (muL, muR, varL, and varR) to reference sample locality mean and variance values. I limited the amount of geographic space explored by the MCMC (-30, 180 km). Because of the large number of free variables for model II (n = 7), I reduced the tune parameter from 1.5 to 1.2. MCMC chains were set to 100,000 with a burnin of 10,000. Convergence was assed by visually inspecting the MCMC traces for each parameter. I considered the model with the lowest AICc score as the best-fitting model. To compare among parameter estimates for different traits, I considered estimates that have non-overlapping confidence intervals to be significantly different, following methods in (Ruegg 2008).

*Genetic Admixture*

I fit a single cline (sexes pooled) to admixture assignments based on the STRUCTURE analysis of informative SNPs (n = 23). Because these assignments are continuous and numerical in nature, I used the same five quantitative trait models used for the bill color and bill width measurements. I used the same settings for geographic space explored (-30, 180 km), tune parameters (1.2 for Model II), and length of MCMC chains (100,000 discarding 10,000 as burnin). As for the bill color and width clines, I checked convergence by visually assessing the traces for individual parameters, and I

performed model selection by considering the model with the lowest AICc score as the best-fitting model.

*Neutral Expectations*

I used Endler's neutral diffusion equation (1977) to calculate neutral expectations for cline widths.

$$T = .35\ (w/d)^2$$

Where T = time since contact, *w* is cline width (1/max slope), and d = dispersal distance (standard deviation of parent - offspring distances). Natal dispersal distances are not well known in hummingbirds. However, they are generally strong flyers. Daily movements can often be nearly one kilometer or even larger for some hummingbird species (Gary Stiles, pers. comm.) Dispersal distances for hummingbirds may be dozens of times larger than their daily movements, similar to other land bird (Eric Johnson, pers. comm.). Paradis et al. (1998) report 14 km natal dispersal for the barn swallow, *Hirundo rustica.* I calculated Endler's equation using a range of dispersal distances smaller than, equal to, and greater than the known distance for barn swallow: 1 km, 14 km, and 50 km.

**RESULTS**

*Phenotypic variation*

The bill color index was significantly different between male *T. polytmus* and *T. scitulus* individuals ($X^2$ = 35, df = 2, *p*-value < 2.51 E -8) and female *T. polytmus* and *T. scitulus* individuals ($X^2$ = 22, df = 2, *p*-value = 1.67 E -05) (Table S3.3 A). The ANOVA for bill width variation was highly significant (*p*-value = 2 E -16; Table S3.3 B). The post-hoc Tukey Test (Table S3.3 C) procedure revealed highly significant differences for several comparisons: 1) male *T. polytmus* and *T. scitulus* (adjusted *p*-value = 0), 2) male

and female *T. polytmus* (adjusted *p*-value = 1.26 E -6), and 3) female *T. polytmus* and *T. scitulus* (adjusted *p*-value = 5.60 E -5). No significant difference in bill width was found between male and female *T. scitulus* (adjusted *p*-value = 0.55).

Little evidence for observer bias was found in measurements of the three skeletal elements used in this study: sternum length, keel length, and keel depth. The percent coefficient of variation for the inter-observer treatment was similar to, or lower than, the intra-observer treatment for these three elements (Fig. S3.3). A PCA of all specimens did not reveal any structure or artificial grouping by observer (Fig. S3.4). I found no evidence for interspecific body size differences in males (*p*-value = 0.54) or females (*p*-value = 1.00). But, strong intra-specific differences in body size were detected for both species (Table S3.4).

*Microsatellites*

The microsatellite loci Tro21, Tro18, and Tro13 were out of Hardy-Weinberg equilibrium (*p*-value < 0.001). These same loci showed strong evidence of possessing high frequencies of null alleles ($0.20 \leq f_o \leq 0.46$). Because the presence of null alleles can bias estimates of population differentiation, these loci were excluded from further analysis. Of the six remaining loci, Tro19 and Tro6 showed slight departures from Hardy-Weinberg equilibrium (*p*-value = 0.02, 0.04, respectively), but none showed evidence of linkage disequilibrium (*p*-value > 0.003, adjusted at 5% nominal level). All had low estimated frequencies of null alleles ($f_o \leq 0.08$).

I detected 55 alleles in 171 individuals in six microsatellite loci (Tro2, Tro3, Tro5, Tro6, Tro17, Tro19) included in the study. The number of alleles per locus ranged from 6 to 18. Five alleles were private to Site 8 (*n* = 38), one allele private to sample Site 3 (*n* =

12), and one allele private to sample Site 5 (n = 18). Per-locus estimates of gene diversity and allelic richness were high at most sampling localities (Table S3.5). A comparison between reference *T. polytmus* and *T. scitulus* sites revealed that *T. polytmus* did not have higher estimates for gene diversity or allelic richness than *T. scitulus* (one-sided *p*-values based on 1,000 permutations >> 0.05).

Tests of differentiation showed low divergence in pairwise comparisons between sites (Table 3.2). The estimate of $F_{ST}$ over all microsatellite loci and sample localities was 0.032. While small, the $F_{ST}$ estimate was significantly different from zero (*p*-value = 0.001; CI = [0.008, 0.074]). However, most pairwise $F_{ST}$ values were statistically indistinguishable from zero. Of the 21 pairwise comparisons between the parental reference sites, only two comparisons showed weak differentiation. Site 8, which is located in the hybrid zone, showed weak differentiation with five other sites: two *T. polytmus* sites, two hybrid zone sites, and one *T. scitulus* site. The AMOVA showed that 95% of the genetic variation was within sites. Though small, the among-site component of genetic variation was significant ($\Phi_{PT}$ = 0.051, P = 0.001; Table S3.6), but there was no obvious geographic pattern to this variation.

Consistent with the $F_{ST}$ estimates and AMOVA, I found no detectable geographic structuring of the microsatellite data based on a Bayesian clustering analysis in STRUCTURE. The posterior probability for K = 1 was the least negative of all posterior probabilities for K = 1 – 5. The posterior probability of assignment to each genetic cluster (K = 2) was approximately 50% for all individuals, indicating there is no signal for genetic divergence at these loci (Fig. 3.2 A).

Table 3.2. Estimates of genetic differentiation based on 6 microsatellite loci for 171 individuals across 12 sites of *T. polytmus*, *T. scitulus*, and putative hybrids. Pairwise $F_{ST}$ values are above the diagonal, and corresponding significance levels are below. Strict Bonferroni-corrected *p*-value at $\alpha = 0.05$ is .000758. NS = not significant, * = significance at the 5% nominal level.

| | Site 1 | Site 2 | Site 3 | Site 4 | Site 5 | Site 6 | Site 7 | Site 8 | Site 9 | Site 10 | Site 11 | Site 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Site 1 | | 0.013 | -0.001 | 0.004 | 0.010 | -0.006 | 0.005 | 0.036 | 0.045 | 0.040 | 0.063 | -0.011 |
| Site 2 | NS | | 0.000 | 0.039 | -0.008 | 0.011 | 0.017 | 0.038 | 0.041 | 0.026 | 0.089 | 0.020 |
| Site 3 | NS | NS | | 0.009 | 0.016 | 0.004 | 0.007 | 0.038 | 0.026 | 0.029 | 0.084 | -0.008 |
| Site 4 | NS | NS | NS | | 0.043 | 0.026 | 0.022 | 0.054 | 0.061 | 0.058 | 0.073 | 0.008 |
| Site 5 | NS | NS | NS | NS | | -0.004 | 0.010 | 0.042 | 0.045 | 0.059 | 0.087 | 0.033 |
| Site 6 | NS | NS | NS | NS | NS | | -0.004 | 0.046 | 0.030 | 0.069 | 0.086 | -0.012 |
| Site 7 | NS | NS | NS | NS | NS | NS | | 0.053 | 0.032 | 0.051 | 0.095 | 0.006 |
| Site 8 | * | NS | * | NS | NS | * | * | | 0.012 | 0.008 | 0.005 | 0.034 |
| Site 9 | NS | NS | NS | NS | NS | NS | NS | NS | | 0.014 | 0.038 | 0.035 |
| Site 10 | NS | NS | NS | NS | NS | * | NS | NS | NS | | 0.032 | 0.053 |
| Site 11 | * | NS | * | NS | NS | * | * | NS | NS | NS | | 0.065 |
| Site 12 | NS | NS | NS | NS | NS | NS | NS | * | NS | NS | * | |

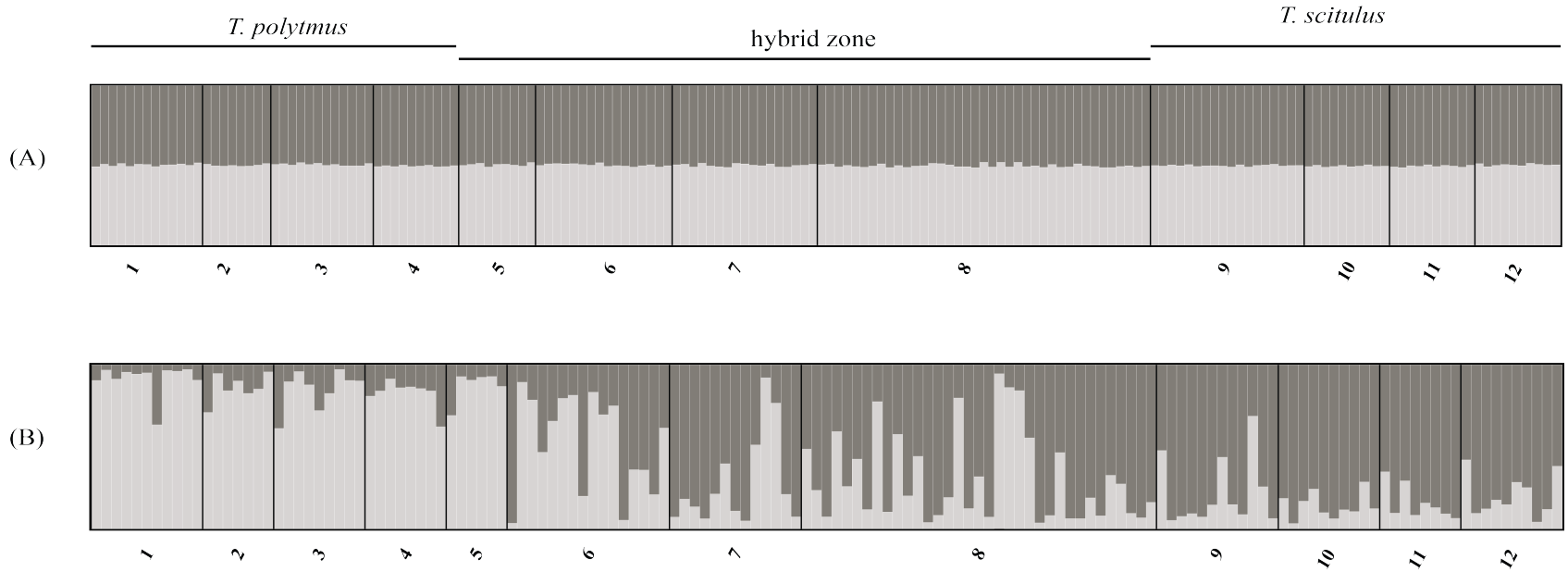



Fig. 3.2. (A) STRUCTURE admixture proportions (K = 2) for six microsatellite markers in 172 individual *T. polytmus*, *T. scitulus*, and their hybrids across 12 sites. (B) STRUCTURE admixture proportions (K = 2) for 23 "informative" SNPs in 145 individual *T. polytmus*, *T. scitulus*, and their hybrids across 12 sites.

*Genotyping-by-sequencing*

The FASTSTRUCTURE analysis performed on all SNPs (n = 6,451) did not reveal any interpretable pattern of genetic structure; all individuals assigned strongly to one genetic cluster (Fig. S3.5). The PCA revealed modest but obvious separation between species, with most phenotypic hybrids falling in the middle (Fig. 3.3).

The discriminant analysis of principal components (DAPC) based on 30 retained principal components showed even stronger, though imperfect discrimination among the three morpho-groups (Fig. 3.4). The densities plot for the first discriminant function reveals at least some overlap between parental species (Fig. 3.5). For the subset of informative SNPs that had loadings above the threshold, allele frequency differences between parental reference sites ranged from 0 to 0.6, of which only five loci had allele frequency differences greater than 0.5. The locus with the highest loading also had the highest absolute allele frequency difference between parental reference sites for both alleles: 0.6. No SNP had fixed or nearly fixed differences ($0.2 < f_0 < 0.8$) between reference populations (Table S3.7).

The STRUCTURE analysis performed on just the informative SNPs (n=23) resolved two distinct genetic clusters (Fig. 3.2 B). Higher values of K did not capture additional structures in the data (results not shown). The structure plot shows strong assignments to one genetic cluster by *T. polytmus* individuals in Sites 1-3, strong assignment to the other genetic cluster by *T. scitulus* individuals in Sites 9-12, and a range of assignments in hybrid populations.

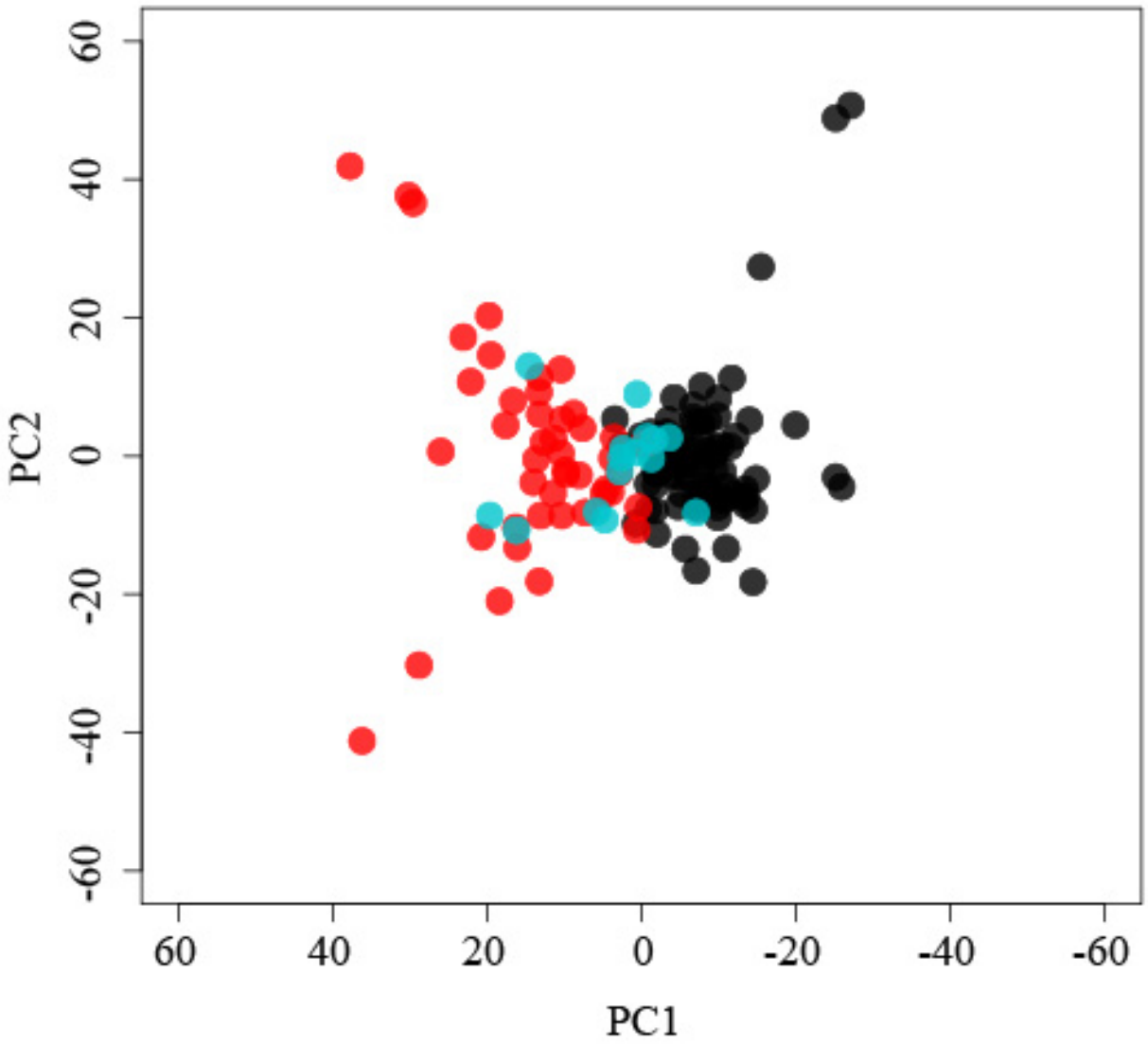



Fig. 3.3. Principal component analysis (PCA) of 145 individual *T. polytmus*, *T. scitulus*, and their hybrids in ordinate space. *Trochilus scitulus* = black circles, *T. polytmus* = red circles, and phenotypic hybrids = teal individuals.

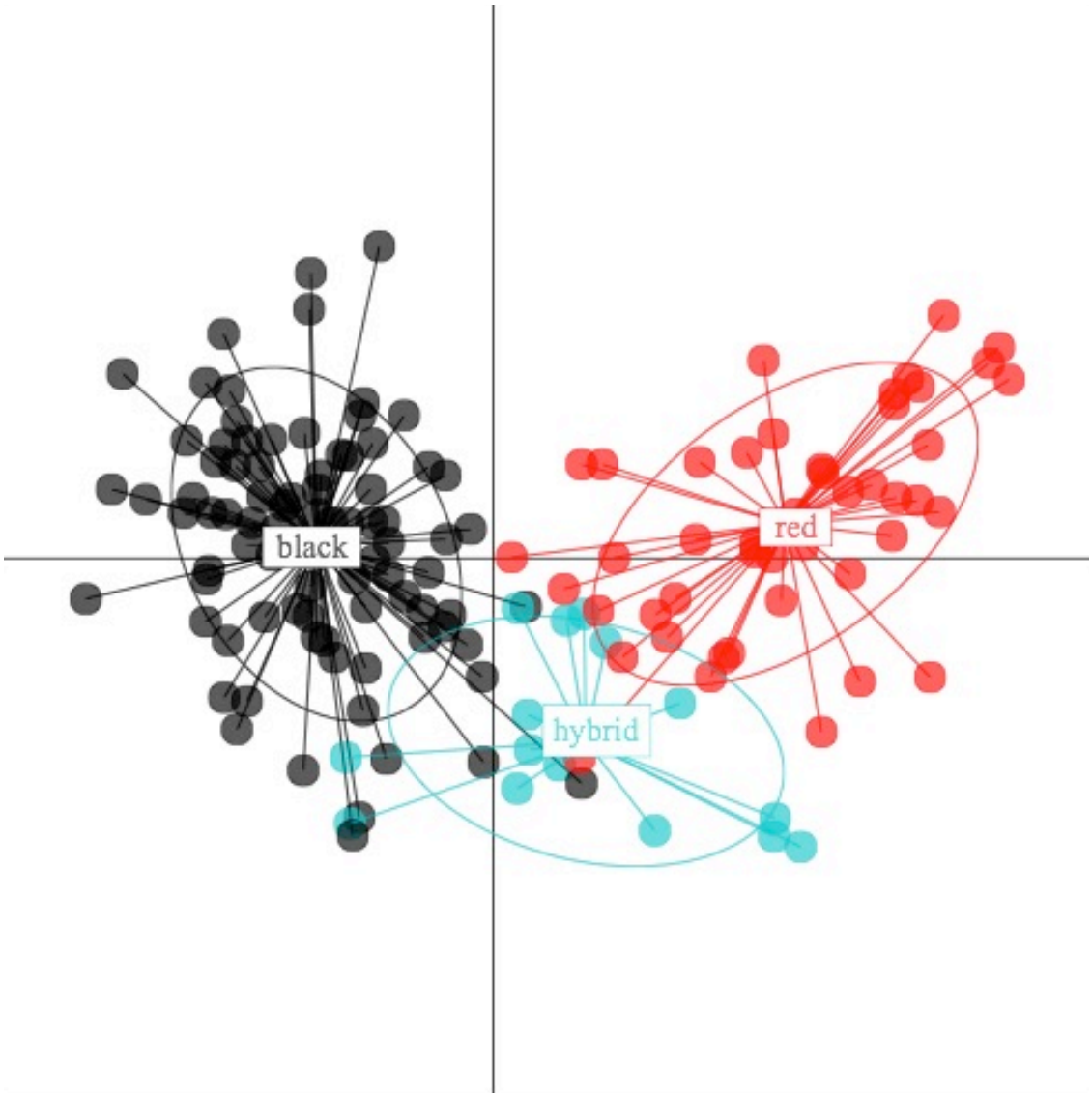



Fig. 3.4. Discriminant analysis of principal components (DAPC) based on 30 retained principal components in 6,451 SNPs and 145 individual *T. polytmus*, *T. scitulus*, and hybrids.

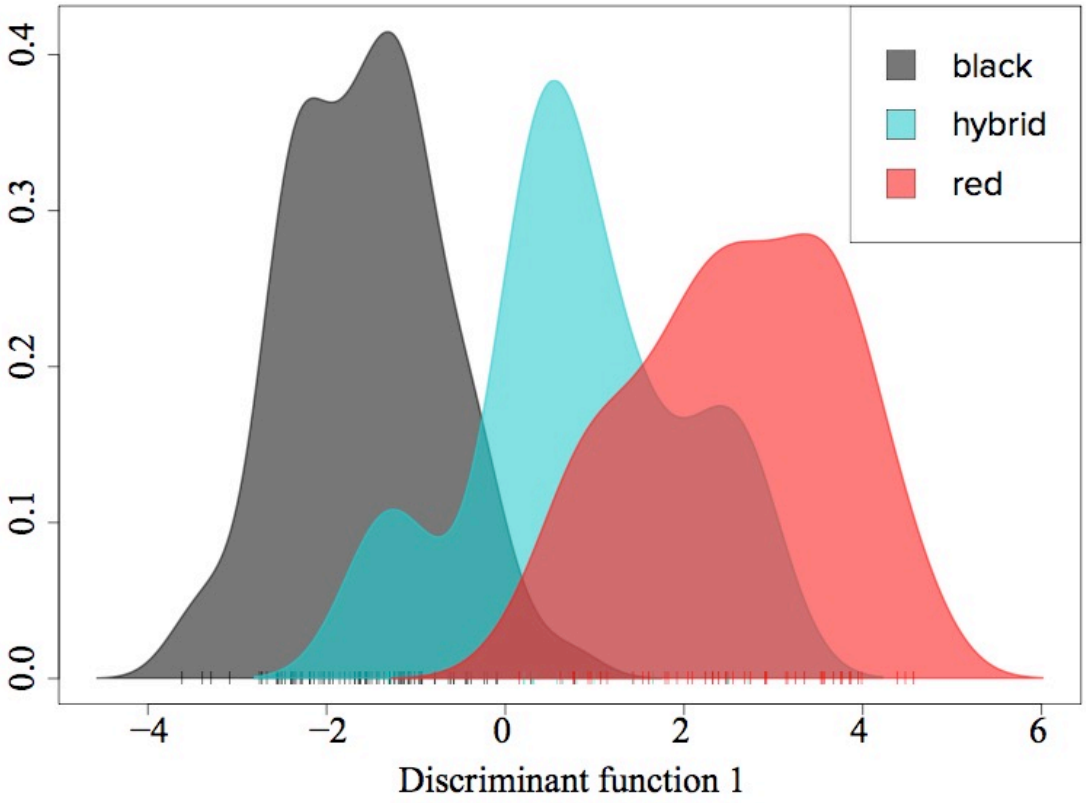



Table 3.5. Density of individuals along the first discriminant function. Red corresponds to *T. polytmus* individuals, black corresponds to *T. scitulus* individuals, and teal corresponds to hybrid individuals.

*Geographic cline analysis*

Clines for bill color were shockingly narrow for both sexes: only 2.2 km for males and 3.74 km in females (Fig. 3.6 A-C). Cline models for bill width were somewhat wider in both males (13.9 km) and females (16.0 km). Male clines were significantly narrower than female clines for both traits examined, and the cline centers for all morphological clines were statistically indistinguishable. Additionally, estimated variance was higher at the center of the clines than in the tails for all traits (bill color, bill width, and genetic admixture) (Table 3.3).

The cline for genetic admixture (sexes pooled) was also very narrow: 6.75 km (Fig. 3.7). Interestingly, its geographic cline center was identical to that for male bill width (135.8 km from Site 1), and its estimated variance was also higher at the center of the cline than in either tail. All cline centers for phenotypic and genetic admixture traits were coincident with the Rio Grande Valley in Eastern Jamaica.

Table 3.3. Estimates of genetic differentiation based on 6 microsatellite loci for 171 individuals across 12 sites of *T. polytmus*, *T. scitulus*, and putative hybrids. Pairwise $F_{ST}$ values are above the diagonal, and corresponding significance levels are below. Strict Bonferroni-corrected *p*-value at $\alpha = 0.05$ is .000758. NS = not significant, * = significance at the 5% nominal level.

| Group | Trait | Best Model | Center (CI) | Width (CI) | varH (CI) |
|---|---|---|---|---|---|
| males | bill color | I | 135.8 (135.1 - 136.3) | 2.2 (0.20 - 4.8) | 2.21 (0.71 - 86,494) |
| males | bill width | I | 132.4 (130.4 - 134.0) | 13.9 (9.2 - 20.8) | 1.8 E -4 (1.5 E -8 - 0.02) |
| females | bill color | V | 134.3 (133.3 - 135.0) | 3.74 (0.84 - 4.50) | 1.17 (1.13 - 7.03) |
| females | bill width | I | 135.3 (133.3 - 137.6) | 16.0 (3.12 - 29.8) | 2.39 E -5 (9.14 E -7 - 0.02) |
| sexes pooled | admixture | I | 135.8 (135.0 - 136.5) | 6.75 (4.70 - 9.16) | 0.04 (0.03 - 0.06) |

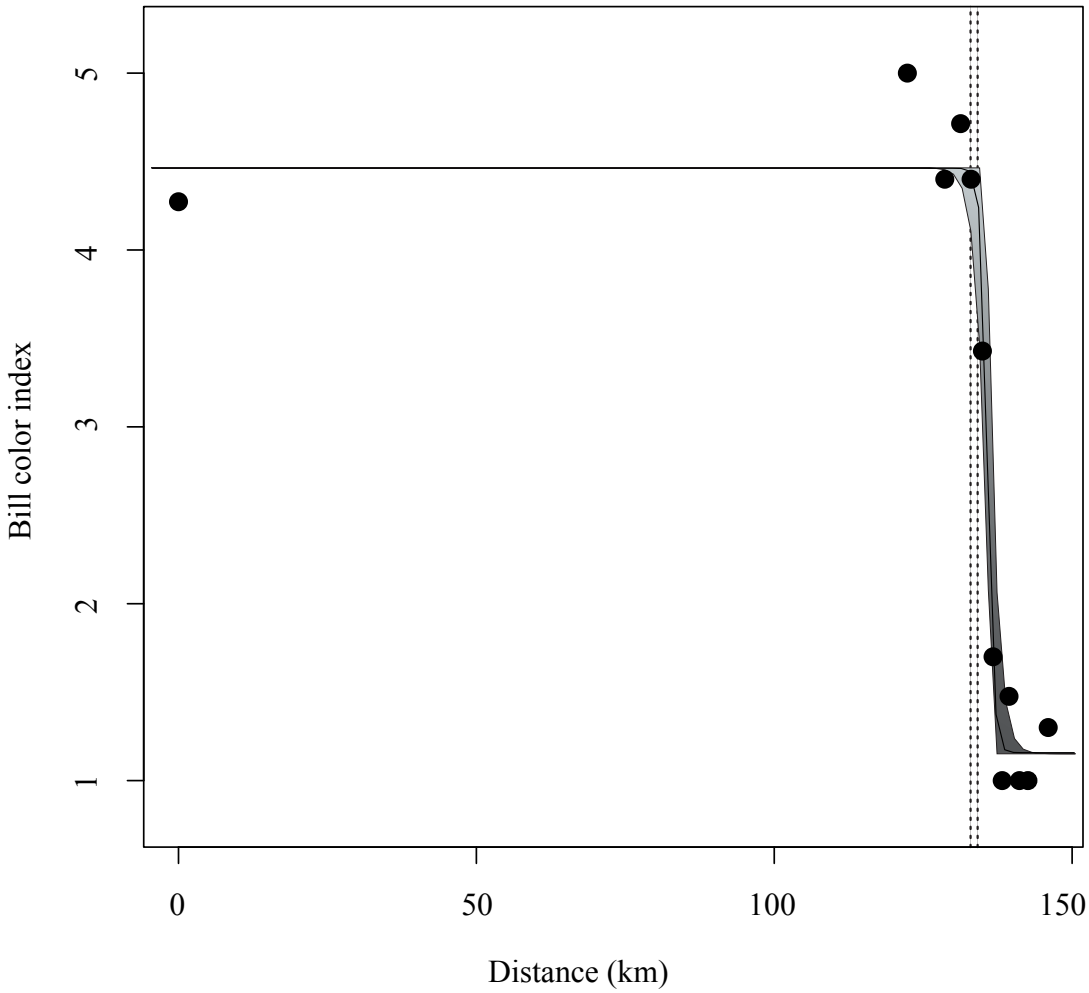



Fig. 3.6 A. Cline for male bill color. Gray shading indicates the 95% confidence intervals, and the vertical dotted lines represent the east and west boundaries of the Rio Grande River.

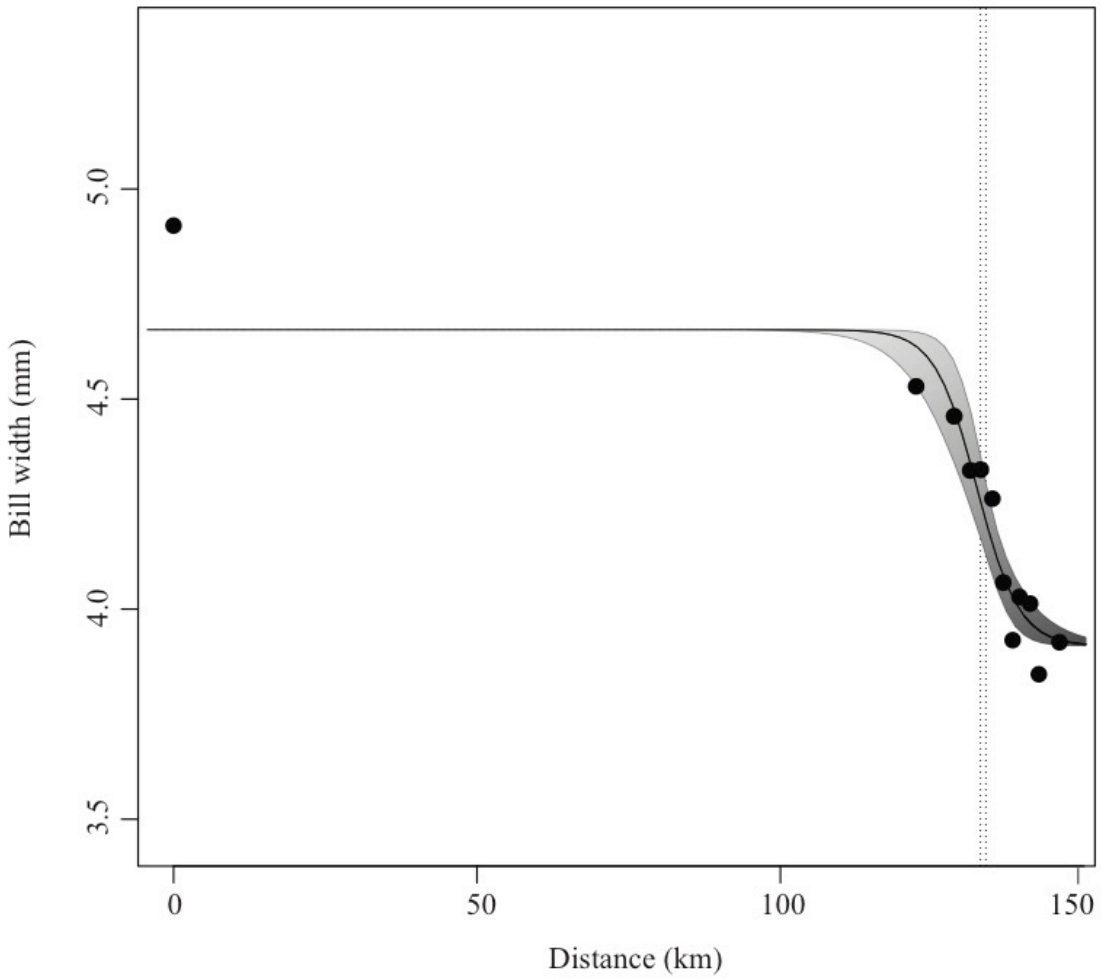



Fig. 3.6 B. Cline for male bill width. Gray shading indicates the 95% confidence intervals, and the vertical dotted lines represent the east and west boundaries of the Rio Grande River.

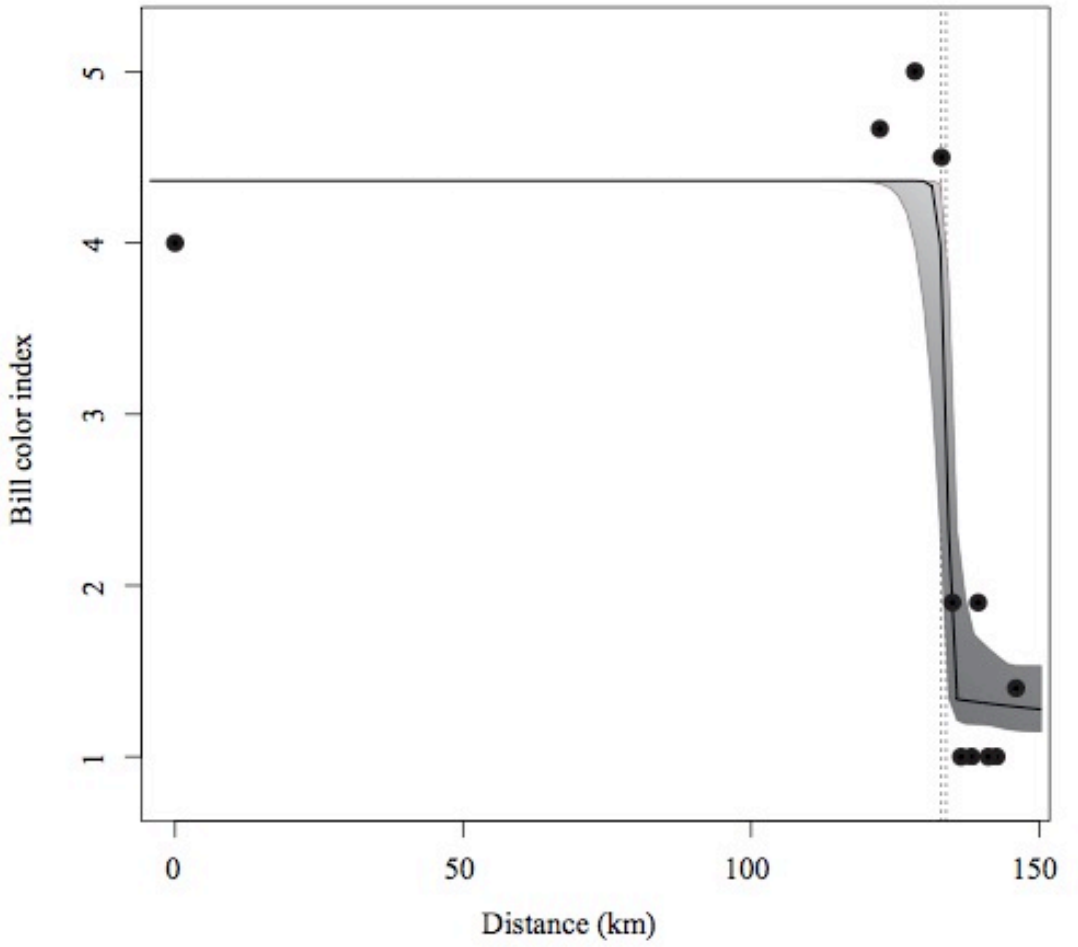



Fig. 3.6 C. Cline for female bill color. Gray shading indicates the 95% confidence intervals, and the vertical dotted lines represent the east and west boundaries of the Rio Grande River.

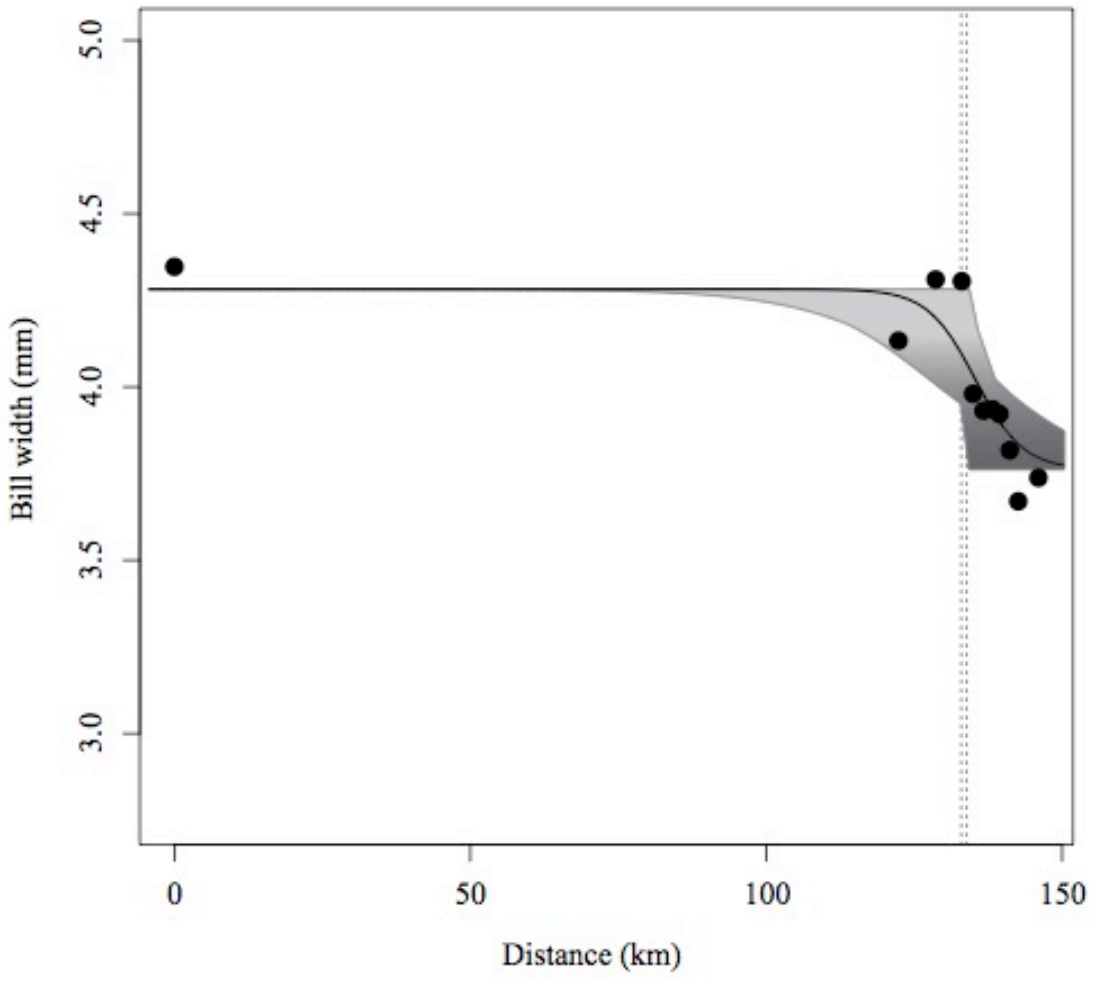



Fig. 3.6 D. Cline for female bill width. Gray shading indicates the 95% confidence intervals, and the vertical dotted lines represent the east and west boundaries of the Rio Grande River.

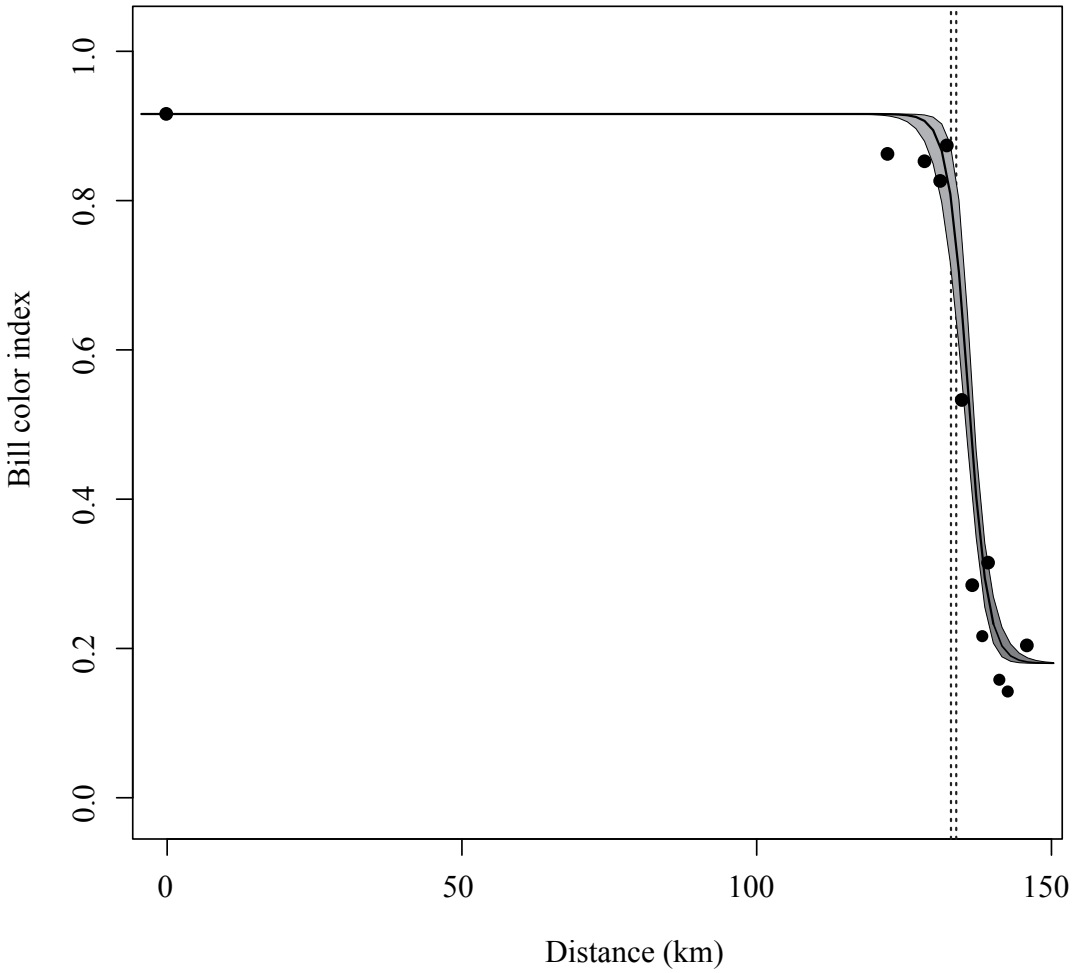



Fig. 3.7 Cline for admixture assignments (sexes pooled). Gray shading indicates the 95% confidence intervals, and the vertical dotted lines represent the east and west boundaries of the Rio Grande River.

## DISCUSSION

Parapatric streamertail hummingbirds (*Trochilus polytmus* and *T. scitulus*) on the island of Jamaica are divergent in bill color and bill width, with both bill traits changing rapidly across the narrow hybrid zone. Along the same transect we detected genomic differentiation in a 6,451-SNP dataset. However, the amount of genomic differentiation was extremely low. For example, six microsatellite loci, markers that are typically

capable of detecting weak differentiation (Bruford and Wayne 1993) were unable to distinguish the taxa. Neither did a discriminant analysis of principal components (DAPC) of 6,451 genome-wide SNPs fully discriminate the species. The DAPC approach is exceptionally powerful for discriminating closely related organisms (Jombart et al. 2010), so it is all the more surprising that this analysis failed to fully discriminate the *Trochilus* parental species.

In the 6,451 SNP dataset, no SNP had fixed or nearly fixed ($0.2 < f > 0.8$) allele frequency differences between reference parental sites. Only one SNP contributed majorly to the DAPC, and the loadings for each of its alleles were at least twice as high as the next highest loading, indicating that the cumulative signal of divergence in the other SNPs was comparably much lower. Nonetheless, Bayesian-derived admixture proportions based on this narrow set of "informative" SNPs revealed cline-like variation among *T. polytmus*, *T. scitulus*, and hybrids, suggesting that some of these 23 SNPs may be in or near gene regions that underpin the observed morphological differences.

Empirical studies of young or closely related avian lineages repeatedly show that strong divergence in sequence or gene expression can occur in a narrow set of genes where background divergence is weak or absent (Toews et al. 2015). For example, Poelstra et al. (2014) attribute phenotypic divergence in carrion and hooded crows to localized selection on pigmentation and visual perception genes in otherwise undifferentiated genomes. Toews et al. (2015) recovered only six divergent genomic peaks (four regions near pigment genes) in a whole-genome resequencing study of blue- and golden-winged warblers. Styrjewski and Sorenson (2017) show recombination of ancestral genetic variation in a very narrow set of genes explains phenotypic patterns in a

rapid radiation of munias. Such disparate signals for phenotypic and neutral genetic divergence occur in non-avian early-divergent taxa as well (e.g. sunflowers: Andrew and Rieseberg 2013; Renaut et al. 2013, butterflies: Nadeau et al. 2014), underscoring how rapidly phenotypic traits can diverge in nature, even in the absence of neutral genetic variation.

*Characteristics of the hybrid zone*

I found exceptionally narrow clines for male bill color (2.2 km) and bill width (13.9 km). Female clines were similar to, though slightly wider than, male clines for each trait. In particular, the narrow cline for male bill color (2.3 km) is among the narrowest clines reported from an avian hybrid zone, similar to the dramatically narrow clines for belly (3.0 km) and collar color (4.4 km) reported for lekking manakins in Panama (Butlin et al. 1991; Brumfield et al. 2001), and head coloration (5.1 km) in Reunion grey white-eyes (Delahaie et al. 2017).

The estimated cline centers for genetic and morphological traits are coincident with the Rio Grande Valley in eastern Jamaica. The Rio Grande is one of eight major rivers in Jamaica and separates the eastern Blue and John Crow Mountains. The geological history of this area is fairly complex. While most of the island is composed of limestone and likely emerged approximately 14 MYA in the mid-Miocene, the region to the east of the valley may have uplifted more recently, less than 2 MYA, during the lower to middle Pleistocene (Mitchell, Simon, University of the West Indies at Mona, pers. comm). It is interesting to speculate whether the uplift of the John Crow Mountains created an opportunity for vicariance, with recent hybridization between the parapatric forms erasing most of that signal for divergence.

Another possibility is that the Rio Grande River itself might be an effective dispersal barrier. The Rio Grande is narrow (~70 m wide) and does not seem like an important barrier for a vagile organism like a hummingbird. However, the agricultural activities in the larger valley, including removal of forest, may effectively increase the size of the perceived river barrier. Small dispersal barriers seem to have played a role in the diversification of lowland forms of the Reunion grey white-eyes (*Zosterops borbonicus*). These recently diverged but morphologically distinct taxa are distributed parapatrically and separated by narrow rivers or lava fields (Gill 1973, Delahaie et al. 2017). Thus, small barriers such as rivers may be effective to limit dispersal and create vicariance even for volant species in lowland island systems.

Regardless of past history, it appears that the valley is playing a role in contemporary patterns. Valleys often represent avian population density troughs. Hybrid zones maintained as tension zones can become geographically "stuck" in such troughs (Barton and Hewitt 1985, 1989). The black-billed streamertail density is known to be very high (> 900 individuals per $km^2$) despite its limited geographic range (Chapter 2). Other field observations suggest that streamertails may have higher population densities in the mountains (Bond 1956). If so, densities of both species may be lower in the valley, especially if agricultural activity has limited their numbers. However, more studies are needed to estimate densities in the valley to determine whether indeed it represents a density trough.

I calculated Endler's neutral diffusion equation (1977) using the estimated cline width for male bill color (2.2 km), a one-kilometer dispersal distance, and a generation time of one year. Under these parameters, the *Trochilus* hybrid zone would have to have

formed less than five years ago. Hybrids were first documented nearly 70 years ago, and the current extent of the hybrid zone has been stable for almost 50 years (Bond 1956; Gill et al. 1973; Graves 2015). Correct inference of neutral diffusion and selection depends on accurate measures of dispersal (Baldassarre et al. 2014), which is unknown in *Trochilus.* Nonetheless, even using this conservative estimate for dispersal, it seems likely that some kind of natural selection is acting to maintain these narrow zones.

In selection-dependent zones, narrow clines can be produced by advancing waves of introgression from one species into another (Ferris et al. 1983). The best evidence against the advancing wave hypothesis is having multiple concordant clines for various traits (Barton and Hewitt 1985). I cannot completely rule out this hypothesis without a more detailed understanding of the underlying genetic architecture of bill color and bill width, since these traits could be linked in some way, as could be the SNPs. However, several lines of evidence point to the tension zone model. First, the zone has been stable for 50 generations. Second, the pair shows other subtle differences in morphology. *Trochilus scitulus* has a shorter bill, shorter wings, and a longer, more deeply forked tail than *T. polytmus* (Graves 2015). Additionally, aviary studies show some evidence for behavioral differences in courtship and song (Schuchmann 1978, 1980). These factors point to a tension zone model maintained as a balance between selection against hybrids offset by dispersal into the zone.

Assuming that the streamertail hybrid zone is a tension zone, the strength of selection against hybrid bill color must be extraordinary. Streamertail hummingbirds are polygynous and form leks based on resource defense, factors that increase the strength of sexual and other forms of social competition. Sexual selection accelerates signal

evolution (Seddon et al. 2008) and can drive speciation in the absence of other selective forces (West-Eberhard 1983; Higashi et al. 1999). More work is needed to explore the potential for sexual selection to promote divergence in bill color for streamertail hummingbirds and to understand the role of mate choice in maintaining this narrow zone.

## REFERENCES


Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pp. 267–281 *in* B. Petrov and F. Csaki, eds.

Andrew, R. L., and L. H. Rieseberg. 2013. Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. Evolution 67:2468–2482.

Baldassarre, D. T., T. A. White, J. Karubian, and M. S. Webster. 2014. Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution 68:2644–2657.

Barton, N. H. 1979. The dynamics of hybrid zones. Heredity 43:341–359.

Barton, N. H., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. Nature 341:497–503.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Bleiweiss, R. 1998. Slow rate of molecular evolution in high-elevation hummingbirds. Proc. Natl. Acad. Sci. U. S. A. 95:612–616.

Bond, J. 1956. Check-list of the birds of the West Indies. Academy of Natural Sciences of Philadelphia, Philadelphia.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635.

Brewster, W., and O. Bangs. 1901. On an overlooked species of *Aithurus*. Proc. N. Engl. Zoöl. Club 2:47–50.

Bruford, M. W., and R. K. Wayne. 1993. Microsatellites and their application to population genetic studies. Curr. Opin. Genet. Dev. 3:939–943.

Brumfield, R. T., R. W. Jernigan, D. B. McDonald, and M. J. Braun. 2001. Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. Evolution 55:2070–2087.

Butlin, R. K., M. G. Ritchie, and G. M. Hewitt. 1991. Comparisons among morphological characters and between localities in the *Chorthippus parallelus* hybrid zone (Orthoptera: Acrididae). Philos. Trans. Biol. Sci. 334.

Coyne, J. A., and T. D. Price. 2000. Little evidence for sympatric speciation in island birds. Evolution 54:2166–2171.

Delahaie, B., J. Cornuault, C. Masson, J. a. M. Bertrand, Y. X. C. Bourgeois, B. Milá, and C. Thébaud. 2017. Narrow hybrid zones in spite of very low population differentiation in neutral markers in an island bird species complex. J. Evol. Biol. 12:2132–2145.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. HZAR: hybrid zone analysis using an R software package. Mol. Ecol. Resour. 14:652–663.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A Robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379.

Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Ferris, S. D., R. D. Sage, C. M. Huang, J. T. Nielsen, U. Ritte, and A. C. Wilson. 1983. Flow of mitochondrial DNA across a species boundary. Proc. Natl. Acad. Sci. U. S. A. 80:2290–2294.

Gill, F. B., F. J. Stokes, and C. Stokes. 1973. Contact zones and hybridization in the Jamaican hummingbird, *Trochilus polytmus* (L.). The Condor 75:170–176.

Gosse, P. H. 1847. The birds of Jamaica. Van Vorst, London.

Goudet. 1995. Fstat version 1.2: a computer program to calculate F statistics. J. Hered. 86:485–486.

Goudet, J., M. Raymond, T. de Meeüs, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144:1933–1940.

Graves, G. R. 2015. A primer on the hybrid zone of Jamaican streamertail hummingbirds (Trochilidae: *Trochilus*). Proc. Biol. Soc. Wash. 128:111–124.

Graves, G. R. 2009a. Ontogeny of bill color in male streamertail hummingbirds (*Trochilus*). J. Caribb. Ornithol. 44–47.

Graves, G. R. 2009b. Skeletal correlates of body weight in the black-billed streamertail (*Trochilus scitulus*) of Jamaica. Caribb. J. Sci. 45:69–72.

Haldane, J. B. S. 1948. The theory of a cline. J. Genet. 48:277–284.

Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University Press.

Hewitt, G. M. 1988. Hybrid zones–natural laboratories for evolutionary studies. Trends Ecol. Evol. 3:158–167.

Higashi, M., G. Takimoto, and N. Yamamura. 1999. Sympatric speciation by sexual selection. Nature 402:523–526.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components:a new method for the analysis of genetically. BMC Genet. 11:94.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16:1099–1106.

Key, K. H. L. 1968. The concept of stasipatric speciation. Syst. Zool. 17:14–22.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15:1179–1191.

Lance, S. L., C. Hagen, T. C. Glenn, R. T. Brumfield, K. F. Stryjewski, and G. R. Graves. 2009. Fifteen polymorphic microsatellite loci from Jamaican streamertail hummingbirds (*Trochilus*). Conserv. Genet. 10:1195–1198.

Larson, E. L., C. Guilherme Becker, E. R. Bondra, and R. G. Harrison. 2013. Structure of a mosaic hybrid zone between the field crickets *Gryllus firmus* and *pennsylvanicus*. Ecol. Evol. 3:985–1002.

McGuire, J. A., C. C. Witt, J. V. Remsen, A. Corl, D. L. Rabosky, D. L. Altshuler, and R. Dudley. 2014. Molecular phylogenetics and the diversification of hummingbirds. Curr. Biol. 24:910–916.

Moore, W. S. 1977. An evaluation of narrow hybrid zones in vertebrates. Q. Rev. Biol. 52:263–277.

MacColl, A., and S. Lewis. 2000. Hybridisation and Ecology of Jamaican Streamertail Hummingbirds. BirdLife Jamaica 4–10.

Mousadik, A. E., and R. J. Petit. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor. Appl. Genet. 92:832–839.

Nadeau, N. J., M. Ruiz, P. Salazar, B. Counterman, J. A. Medina, H. Ortiz-Zuazaga, A. Morrison, W. O. McMillan, C. D. Jiggins, and R. Papa. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. Genome Res. 24:1316–1333.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U. S. A. 70:3321–3323.

Ortiz-Crespo, F. 1972. A new method to separate immature and adult hummingbirds. The Auk 89:851–857.

Paradis, E., S. R. Baillie, W. J. Sutherland, and R. D. Gregory. 1998. Patterns of natal and breeding dispersal in birds. Journal of Animal Ecology 67:518–536.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295.

Petit, R. J., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. Conserv. Biol. 12:844–855.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Muller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:155.

Raj, A., M. Stephens, and J. K. Pritchard. 2014. FASTSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics 197:573–589.

Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, J. E. Bowers, J. M. Burke, and L. H. Rieseberg. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Commun. 4:1827.

Rohlf, F. J. 2016. tpsDIG, Version 2.10. Dep. Ecol. Evol. State University of New York, Stony Brook NY.

Rosenberg, N. A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138.

Ruegg, K. 2008. Genetic, morphological, and ecological characterization of a hybrid zone that spans a migratory divide. Evolution 62:452–466.

Schuchmann, K. L. 1978. Allopatrische artbildung dei der kolibrigattung *Trochilus*. Ardea 66:156–172.

Schuchmann, K. L. 1980. Die Jamaika kolibris *Trochilus polytmus* und *Trochilus scitulus*. Biotropic-Verlag, Frankfurt am Main.

Seddon, N., R. M. Merrill, and J. A. Tobias. 2008. Sexually selected traits predict patterns of species richness in a diverse clade of suboscine birds. Am. Nat. 171(5):620-631.

Stryjewski, K. F., and M. D. Sorenson. 2017. Mosaic genome evolution in a recent and rapid avian radiation. Nat. Ecol. Evol. 1:1912–1922.

Szymura, J. M., and N. H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. Evolution 40:1141.

Toews, D. P. L., L. Campagna, S. A. Taylor, C. N. Balakrishnan, D. T. Baldassarre, P. E. Deane-Coe, M. G. Harvey, D. M. Hooper, D. E. Irwin, C. D. Judy, N. A. Mason, J. E. McCormack, K. G. McCracken, C. H. Oliveros, R. J. Safran, E. S. C. Scordato, K. F. Stryjewski, A. Tigano, J. A. C. Uy, and B. M. Winger. 2015. Genomic approaches to understanding population divergence and speciation in birds. The Auk 133:13–30.

Torkamaneh, D., J. Laroche, and F. Belzile. 2016. Genome-wide SNP calling from genotyping by sequencing (GBS) data: a comparison of seven pipelines and two sequencing technologies. PloS One 11:e0161333.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22:787–798.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

West-Eberhard, M.J. 1983. Sexual selection, social competition, and speciation. Q. Rev. Biol. 58:155–183.

White, T. A., S. E. Perkins, G. Heckel, and J. B. Searle. 2013. Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Mol. Ecol. 22:2971–2985.

Zusi, Richard L. 2013. Introduction to the skeleton of hummingbirds (Aves: Apodiformes, Trochilidae) in Functional and Phylogenetic Contexts. Pp. 1–94 *in* Ornithological Monographs No. 77. American Ornithologists' Union.

# CHAPTER 4.

# A TRANSCRIPTOMICS APPROACH TO IDENTIFYING THE GENETIC BASIS OF DIVERGENT BILL COLOR IN JAMAICAN STREAMERTAIL HUMMINGBIRDS

## INTRODUCTION

Speciation is increasingly viewed as a continuous process (Powell et al. 2013; Seehausen et al. 2014) in which reproductive isolation is seen as the by-product of selection causing other phenotypes to diverge (Harrison 1998). Thus, a key goal for evolutionary studies is to identify the genetic basis of diverging phenotypes in species at various points along the speciation continuum (Nosil and Feder 2013). Young lineages at early stages in the continuum are useful models for examining the potential for a causal link between the divergent phenotypes and cessation of gene flow (Shaw and Mullen 2011; Nosil and Feder 2013; Powell et al. 2013; Seehausen et al. 2014).

Recent advances in sequencing technologies have opened up new avenues for speciation research in avian taxa (Toews et al. 2015). For example, reduced-representation approaches such as genotyping by sequencing (GBS) and double digest restriction-associated DNA sequencing (ddRADSeq) outperform traditional loci to resolve population structure in species with shallow divergence (e.g. Campagna et al. 2015). However, these data tend to be scattered at sites across the genome, including non-coding sites; therefore, it is sometimes impossible to obtain functional information of the variant containing sequences unless sites in question are tightly linked to particular genes, or a reference genome is available (de Wit et al. 2012).

Revolutionary new approaches to RNA sequencing (RNA-seq) have enabled sequencing of the coding genes in nonmodel organisms. Annotation tools allow for

reliable mapping and annotation of these coding genes (Jax et al. 2018). Thus, RNA-seq experiments have made it possible to examine gene expression differences that underpin divergent phenotypes within a species complex (e.g. Mason and Taylor 2015) or between young lineages at early stages of speciation (Wolf et al. 2010; Uebbing et al. 2016), shedding light on the mode of gene expression evolution (Gilad et al. 2006) and the role of selection to either promote or constrain expression divergence (Whitehead and Crawford 2006).

Birds are remarkably diverse in color traits relative to other vertebrate groups, especially mammals (Galván and Solano 2016). For birds, pigmentation is largely attributed to melanins, a ubiquitous class of polymers that serve multiple functions in avian and other vertebrate taxa. For example, melanin-based traits serve as visual cues in courtship, lend structural support against mechanical damage or feather-degrading bacteria, and confer crypsis (Galván and Solano 2016). Understanding form and function of melanin-based traits and the factors to promote their remarkable diversity has long captivated evolutionary biologists (Darwin 1871; Mayr 1942). More recently, evolutionary biologists have applied genomic tools to identify the genetic basis for divergent melanin-based traits and investigate causal links to speciation (Uy et al. 2018).

Two of the most widely studied genes across birds and other non-avian vertebrates are the melanocortin-1 receptor (*MC1R*) and agouti signaling protein (*ASIP*) (Hubbard et al. 2010). In particular, *MC1R* is highly conserved across vertebrates (Mundy 2005). A single amino acid change has been perfectly linked with plumage variation across island populations of *Monarcha castaneoventris* (Uy et al. 2016), red-footed boobies (Baião et al. 2007), and bananaquits (Theron et al. 2001), among others

(reviewed in Mundy 2005). The *ASIP* gene is a paracrine signaling protein that causes melanocytes to switch from producing eumelanin (dark melanins) to pheomelanin (red and brown melanins) and has known pleiotropic effects as well as epistatic relationships with *MC1R* (Hoekstra 2006; Hubbard et al. 2010). However, melanin-related pathways involve hundreds of genes (Poelstra et al. 2013; Galván and Solano 2016). For example, research involving laboratory mice has uncovered melanin pathway genes including upstream signaling receptors, in addition to their immediate regulators (such as *ASIP*), downstream elements that are directly involved in eumelanin synthesis, and a suite of genes acting during melanosome biogenesis or transport that also have major phenotypic effects (reviewed in Poelstra et al. 2013). More work is needed to characterize the molecular function and phenotypic effects of these lesser known pigmentation genes to better understand causal links to avian speciation (Hubbard et al. 2010).

The distinctive hummingbird genus *Trochilus*, which is represented exclusively by two sexually dichromatic taxa endemic to the oceanic island of Jamaica, has long represented a possible exception to the rule that *in situ* speciation on small oceanic islands does not occur in birds (Poelstra et al. 2013; Galván and Solano 2016). The males of these species are distinguished primarily by bill color (Brewster and Bangs 1901), a secondary sexual ornament, which is coral red and narrowly tipped in black (partially melanized) in the red-billed streamertail (*Trochilus polytmus*), and jet black (fully melanized) in the black-billed streamertail (*T. scitulus*). In field-based collections, the red coloration of *T. polytmus* fades rapidly post mortem, suggesting that the bill appears red due to the presence of highly vascularized tissues rather than a red pigment such as a carotenoid or phyomelanin (Graves 2009, 2015). Phylogenetic (McGuire et al. 2014) and

population genetic analyses (Chapter 3) reveal shallow divergence between these taxa, yet strong selection appears to maintain the divergent bill color trait in the face of ongoing hybridization.

Here, I used a transcriptomics approach to characterize gene expression and sequence divergence between *Trochilus polytmus* and *T. scitulus* hummingbirds. I built a *de novo* transcriptome using paired-end cDNA libraries derived from pectoral muscle tissue of 3 *T. polytmus* and 4 *T. scitulus* individuals, and annotated the transcriptome using Blast2Go (Conesa et al. 2005). I further annotated the transcriptome using Basic Local Alignment Search Tool (blastx) searches against a custom database containing only candidate pigmentation genes. Finally, I identified a panel of single nucleotide polymorphisms (SNPs) to examine variation of interspecific sequence polymorphisms. My specific goals are to 1) examine evidence for gene expression and nucleotide sequence divergence, and 2) identify genes that are putatively involved in melanogenesis or melanin transport as a first step in characterizing the genetic basis for the divergent bill color trait.

## METHODS

### *Field methods*

Eight streamertails were mist-netted at least 90 minutes after sunrise between 4 and 13 December 2014. Red-billed streamertails (*Trochilus polytmus*, n = 4) were netted on the property of Windsor Research Station in Trelawny Parish. Black-billed (*T. scitulus*, n = 4) were netted on Ecclesdown Road in Portland Parish (Table 4.1). Netted streamertails were placed individually in soft muslin bags and suspended in shaded locations with good air circulation for a holding time of at least 15 min prior to

euthanization. After the initial holding period, I photographed the bill of each bird in a dorsal and ventral position using a Panasonic Lumix© digital camera. Following euthanasia, the birds were dissected and the tissue samples removed using sterile dissection tools. A separate scissor and forceps were used for each tissue to avoid cross contamination of tissue-specific RNAs.

Tissues were processed in the following sequence: pectoral muscle (current study), whole brain, whole eyes, heart, and liver. All tissue samples were placed in pre-labeled organ-specific 1.5 mL tubes filled with RNA*Later*™ (Ambion), and were scored heavily to ensure even fixation of tissues in the buffer solution. The post mortem interval (PMI; time between euthanization and stabilization of each individual tissue in RNA*Later*™) was recorded in a field notebook (Cheviron et al. 2011). Following recommendations from Ambion, RNA samples were incubated at room temperature for 24 hours, after which the RNA*Later*™ was drained from the tubes, and tissues were patted dry using sterile cotton and returned to the dry tube. Each tube was then submerged in LN2 for transport back to the United States. The tubes were transferred to a -80° C freezer for permanent storage. Once tissues had been placed in RNA*Later*™, the birds were processed as round skins and partial skeletons. I recorded data for each bird in a field notebook, including location and time of capture, soft part coloration, sex and breeding status, time of euthanasia, and post mortem interval (elapsed time between euthanasia and stabilization of tissue on RNA*Later*™).

*RNA-seq library preparation and sequencing*

I performed mRNA extraction using 2-10 mg of pectoral muscle from each individual. Tissue samples were individually placed in sterile petri dishes and weighed

using an Adam PW 124 (max 120g, d=0.0001g) scale. Samples were then homogenized in individual 1.5 ml ribonuclease-free tubes using a power drill and sterile drill bits, and extracted following the 'mini' mRNA extraction protocol as described in the Dynabeads® mRNA DIRECT™ kit (Invitrogen). To reduce the amount of rRNA contamination, I performed the optional second wash using Buffer 'B', and samples were eluted using the low temperature setting, 65-70° C. Each sample was assessed for quality and quantity using the Agilent® Tape Station.

Rapid Genomics© prepared paired-end, individually barcoded cDNA libraries from the eight muscle samples using in-house protocols. Bioanalyzer® traces were generated to assess quality of the amplified libraries. Due to the presence of adapter dimers, Rapid Genomics© performed an additional round of cleanup. A Bioanalyzer® trace on the pooled library showed that the adapter dimer had been effectively removed. Libraries were individually barcoded and pooled for sequencing to 150bp on the Illumina HiSeq3000.

*Transcriptome assembly*

I used de-multiplexed reads from seven libraries representing 3 *T. polytmus* and 4 *T. scitulus* individuals to assemble a *de novo* transcriptome. Prior to assembly, I used TRIM_GALORE! to trim adapter sequences or partial adapter sequences from the 3' ends of the reads. I also used TRIM_GALORE! to filter low-quality reads with a phred score < 20 and short reads less than < 20 bp (Table S4.1). The TRIM_GALORE! option "--paired" maintained the integrity of the paired data by removing read pairs for which one of the two sequences failed to pass filter, and "orphan" reads were not discarded, but stored in a separate file.

I used the AGALMA pipeline (Dunn et al. 2013) to perform the *de novo* assembly on the trimmed and filtered reads. AGALMA calls TRINITY (Haas et al. 2013) for the assembly step, but, unlike a standard TRINITY run, the AGALMA pipeline first performs subassemblies and uses blast functions to remove contigs and associated reads that map to rRNA (Dunn et al. 2013). Removing RNA prior to the assembly step reportedly improves the efficiency of assembly (Dunn et al. 2013). I performed quality checks of the final *de novo* assembly by 1) counting the number of transcripts that assembled into each transcriptome using the 'grep" command ($ grep -c '>' Trinity.fasta), and 2) generating contig statistics using the TRINITY script, TrinityStats.pl. These statistics include the contig N50, the length for which at least 50% of assembled bases are in transcripts of length x. I based the calculation N50 using the longest isoform per 'gene,' which is slightly more indicative of assembly quality than performing the N50 on all assembled transcripts. Isoforms are alternative mRNAs produced by the same gene, but which differ in their transcription start sites, protein-coding sequence, or untranslated regions (UTRs). I also calculated the ExN50, which scales the N50 by expression to give the same statistic but for just the highest expressed genes as given by percentage threshold.

To further evaluate the quality of the assembly, I used BOWTIE, an ultrafast, memory-efficient short read aligner (Langmead et al. 2009), to map the paired reads in each library back to the *de novo* assembly, and the TRINITY perl script (SAM_nameSorted_to_uniq_count_stats.pl) to evaluate how well the reads mapped back to the assembly. Paired reads that mapped to the same contig were considered properly mapped, whereas paired reads that mapped to different contigs were considered improperly mapped. An assembly's quality is generally considered to be good if > 70%

of the reads in a library map back to the assembly in proper pairs. Finally, I evaluated the *de novo* assembly based on the number of fully assembled coding transcripts. To do this, I performed blastx searches of the translated transcripts against a database of protein sequences (SWISSPROT).

*Transcriptome annotation and development of a candidate gene list*

I annotated the *Trochilus* transcriptome by performing searches against the National Center for Biotechnology Information (NCBI) non-redundant protein database (hereafter, the "nr" database) using blastx and a minimum e-value of 1E-6. I used the top hit to each query sequence to perform mapping of gene ontology terms and functional annotation using the program Blast2GO (Conesa et al. 2005).

Poelstra et al. (2013) performed an exhaustive search of melanin-related genes to identify the genetic basis for plumage polymorphism in carrion and hooded crows. They identified 95 genes. I expanded this list by searching for avian gene and gene products associated with GO terms that had matches to key word descriptions "melanin" and "melanosome" using the AMIGO2 browser, a web-based tool for gene ontology searching (The Gene Ontology Consortium 2015). I discovered 51 GO terms in my search (Table S4.5). I then used these melanin-associated GO terms as queries in AGBASE to identify associated gene and gene products. I discovered 24 of the genes identified by Poelstra et al. (2013). I searched for the remaining 71 genes using Entrez gene symbols in bird genomes listed in Agbase. In some cases, I expanded the search to include all vertebrates to recover well-annotated gene accessions for a given gene symbol. I filtered the results to include only gene and gene products that had corresponding Uniprot accession numbers, and limited the number of representative taxa to no more than five

per gene symbol. In total, I curated 486 accessions representing 198 unique candidate pigmentation genes (Table S4.6). I accessed the sequences of these accessions through UniprotKB, and downloaded them in fasta format, which I then transformed into a custom database. I performed blastx searches against this custom database using a minimum e-value of 1E -6.

*Differential expression*

I compared gene expression profiles among the individuals in this study by aligning each library back to the *de novo* transcriptome. This step was performed using the Trinity wrapper script "align_and_estimate_abundance.pl", which calls the program RSEM (Li and Dewey 2011). The resulting alignments are used to estimate transcript abundance and their credibility intervals. This script outputs a table containing the expression values for each transcript. I then used RSEM to perform cross sample normalization according to the Trimmed Mean of M-values normalization (TMM) method (Robinson and Oshlack 2010). The TMM method is recommended for RNA-seq datasets, for which most (over half) of the genes are not expected to be differentially expressed. The TMM method reports the transcript quantities as Fragments Per Kilobase of cDNA per Million Fragments mapped (FPKM) value for each transcript and individual.

I used EDGER (Robinson et al. 2010) to perform the differential expression analysis at both the gene and isoform level. For pairwise comparisons between two groups, EDGER fits a negative binomial to the data and estimates dispersion using the quantile-adjusted conditional maximum likelihood method. Significance is determined using an exact test that is similar to Fisher's exact test. Because these tests are extremely

powerful, it is easy to detect 'significant differences' no matter how small the difference is. Therefore, a common practice is to consider only the subset of differentially expressed genes that have low adjusted *p*-values and are above a certain log-fold change threshold. Here, I used a FDR < 0.05 and ranked genes in terms of significance above a 4 log-fold change threshold. Isoforms and genes that were identified as being significant were blasted against the NCBI nonredundant protein database BLASTx using a minimum threshold e-value of 1E-4. I used the top hit to each query sequence to perform mapping of gene ontology terms and functional annotation, using the program BLAST2GO (Conesa et al. 2005) to assess the biological function of differentially expressed genes.

*SNP discovery*

I identified a panel of single nucleotide polymorphisms (SNPs) by aligning individual RNAseq libraries to the *de novo* transcriptome. The alignments were performed using the Burrows-Wheeler Aligner (BWA), a read alignment package that efficiently aligns short reads against a reference sequence (Li and Durbin 2009). I used the BWA-mem algorithm, which is recommended for reads over 75 base pairs long. The alignments were outputted in standard SAM (Sequence Alignment / Map) format (Li et al. 2009). Next, I performed a series of pre-processing steps recommended in the GATK best practices for variant calling in RNAseq. These steps include sorting the outputted sam files by coordinates using Samtools, converting the sam to bam format, adding read groups using the Picard tool AddOrReplaceReadGroups, and marking PCR and optical duplicates using the Picard tool MarkDuplicates. Finally, I called SNPs from indexed versions of my bam files using the UnifiedGenotyper tool under default settings (GATK, DePristo et al. 2011). I used the GATK filtering tool VariantFiltration recommended in the

best practices document to filter low-quality SNPs. Specifically, the filter was for Fisher Strand Values greater than 30, a quality score normalized by a read depth minimum of 3, and SNP clusters of at least 3 SNPs within a 35 base pair window. I performed additional filtering for missingness and allelic count. I retained 155,827 bi-allelic SNPs with at least six complete genotypes out of the initial callset of 905,908 SNPs.

*SNP analysis*

Initial patterns of variation in the SNP dataset were visualized using a principal component analysis (PCA) within the R module Adegenet (Jombart 2008). Multivariate approaches such as PCAs are well suited for exploring patterns in large genomic datasets because they can detect structures in the data without making assumptions about the underlying population genetic model (Jombart et al. 2010). I converted a VCF formatted file containing 155,827 filtered RNAseq variants to an Adegenet 'genlight' object using the *vcfR2genlight* function. These SNPs were previously filtered for missingness (see section on SNP discovery, above) and have only 5.67% missing data (at least six of seven genotypes present for each SNP). The class 'genlight' is a formal S4 object for storing genotypes of binary SNPs in a compact way, using a bit-level coding scheme (see R Help Module genlight-class {adegenet} for more information). I performed the PCA on the genlight object containing the 155,827 SNPs using the *glPCA* function under default settings (center=TRUE, scaling=FALSE). Centering is done by subtracting the value from the center of each column from its corresponding column mean (omitting NAs). No scaling was performed on these genetic data.

To more effectively discriminate between species, I performed a discriminant analysis of principal components (DAPC; Jombart et al. 2010). The DAPC yields

synthetic variables of alleles that maximize the between-group variance and minimize within-group variance (Jombart et al. 2010). The DAPC requires a PCA as a prior step, which transforms the data so that variables are uncorrelated and their numbers are less than that of the analyzed individuals, two key assumptions of a discriminant analysis (Jombart et al. 2010). Discriminant analysis of principal components (DAPC) (Jombart et al. 2010) was performed on the first two PCA axes. I conducted two DAPC analyses, one in which I used the *find.clusters* option to identify clusters without *a priori* knowledge of the species, and another in which the two species were defined *a priori*.

I used the program Bayescan to identify candidate SNP loci that may be experiencing positive selection (Foll and Gaggiotti 2008; Foll et al. 2010; Fischer et al. 2011). Bayescan uses allele frequency differences between species or populations to identify $F_{ST}$ outliers. Selection is inferred by decomposing each $F_{ST}$ into a population-specific component (beta), and a locus-specific component (alpha), the latter of which is shared by all populations. Strongly positive values of alpha suggest diversifying selection, whereas negative values suggest balancing selection.

In the context of multiple testing, Bayescan incorporates skepticism about the chance that a given locus is under selection in two ways: by setting the prior odds for the neutral model, and by using the posterior odds (ratio of posterior probabilities for each model), which indicates how probable a model with selection is compared to a neutral model. Bayescan calculates a *q*-value, defined as the false discovery rate analog of the *p*-value, which is the minimum FDR at which this locus may become significant.

To prepare SNPs for analysis in Bayescan, I converted a VCF file containing 155,827 bi-allelic SNPs to Bayescan input format using a custom perl script

("make_bayescan_input.py", author Jason Ladner, available at http://sfg.stanford.edu/; De Wit et al. 2012), setting a quality genotype threshold of 5 and a 2 as the number of individuals required per species. This filtering resulted in a final set of 121,092 bi-allelic SNPs for use in Bayescan. To identify significant outliers, I set the prior odds threshold to 1,000, which is suitable for large datasets with SNPs numbering in the thousands. I compared results to a second analysis where I set the prior odds threshold to 100. In both analyses, I define outliers as loci having the highest posterior odds threshold achieving an FDR of 10%.

Table 4.1. Specimen information including locality name (Windsor = Windsor Research Center in Trelawny Parish, Ecclesdown = Ecclesdown Road in Portland Parish), geographic coordinates rounded to two decimal places (North, West), the elevation at the site of capture in meters over sea level, the date of capture, the initials of preparator (MLB = Matthew L. Brady, CDJ = Caroline D. Judy), and sex of the specimens (M = male). Also included are the time of capture, the time of euthanasia, and the post-mortem interval defined as the elapsed time between euthanasia and subsequent stabilization of muscle tissue in RNAlater©.

| Field ID | Locality | North | West | Elevation (m) | Date | Preparator | Species | Sex | Time of capture | Time of euthanasia | PMI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MLB365 | Windsor | 18.35 | 77.65 | 122 | 5-Dec-14 | MLB | *T. polytmus* | M | 12:30 | 14:25 | 2.0 |
| MLB366 | Windsor | 18.35 | 77.65 | 122 | 5-Dec-14 | MLB | *T. polytmus* | M | 12:30 | 16:45 | 3.0 |
| MLB367 | Windsor | 18.35 | 77.65 | 122 | 6-Dec-14 | MLB | *T. polytmus* | M | 17:30 | 18:13 | 4.0 |
| CDJ238 | Windsor | 18.36 | 77.64 | 100 | 5-Dec-14 | CDJ | *T. polytmus* | M | 10:30 | 13:30 | 10.0 |
| CDJ245 | Ecclesdown | 18.11 | 76.33 | 61 | 9-Dec-14 | CDJ | *T. scitulus* | M | 9:05 | 2:19 | 5.0 |
| CDJ247 | Ecclesdown | 18.11 | 76.33 | 61 | 9-Dec-14 | CDJ | *T. scitulus* | M | 11:30 | 6:01 | 5.0 |
| CDJ249 | Ecclesdown | 18.11 | 76.33 | 61 | 10-Dec-14 | CDJ | *T. scitulus* | M | 8:45 | 11:10 | 5.0 |
| MLB368 | Ecclesdown | 18.11 | 76.33 | 61 | 9-Dec-14 | MLB | *T. scitulus* | M | 8:00 | 9:04 | 2.0 |

## RESULTS

Sample tissues were preserved from 8 individuals (4 *T. polytmus* and 4 *T. scitulus*) with post mortem intervals (PMI) ranging from 2-10 minutes (Table 4.1). The mRNA extractions showed that mRNA was of generally good quality with yields of 14 - 166 nanograms of total mRNA per sample. The multiplexed Hiseq sequencing run produced 28 - 45 million reads per library. The FASTQC reports on the raw library reads (not shown) indicated that the sequencing quality was generally good. Per-base phred scores were greater than 34 in both R1 and R2 across most of the sequence length, with a slightly lower score for the first 6-7 bases. Of note, the quality of the sequence decreased after approximately 100 base pairs. This quality drop at the 5' end of the reads was more pronounced in the R2 reads.

Adapter contamination was significant in all libraries; however, it was highest in *T. polytmus* sample CDJ_238. Quality trimming removed 56% of paired reads from sample library CDJ_238 compared with 3 - 8% of reads for the other libraries (Table S4.1). Therefore, I excluded the sample library for individual CDJ_238 from all downstream analyses. Post-trimming FASTQC reports for the other seven libraries showed that the quality scores improved overall, especially for 5' end of the reads (not shown). The final data set was composed of libraries from three *T. polytmus* and four *T. scitulus*.

*Assembly assessment and annotation*

The final assembly incorporated 181,109,054 bases into 193,916 transcripts, corresponding to 154,348 transcripts, and had an overall GC content of 48.25%. The contig N50 based on all transcripts was 1,360 bp. Considering only the longest isoform per gene, the N50 was 814 (Table S4.2). The N50 statistics scaled by expression (ExN50)

indicated that N50 for the 70% highest expressed genes was 1,587 (Table S4.3). Mapping the individual libraries back onto the assembly resulted in 57 - 71% proper pairs per library, with six of seven libraries close to, but slightly under, the desired 70% threshold (Fig. S4.1). The number of transcripts containing greater than 80% of a protein sequence was 7,426 (Table S4.4).

Searches against the nr database resulted in significant hits to 64,632 transcripts. Of these, 11,846 (~18%) transcripts were mapped and annotated in Blast2Go. The top blast hits were primarily to bird species, with the vast majority to Anna's hummingbird (*Calypte anna*) (Fig. S4.2). The top 20 gene ontology terms mapped by Blast2Go included the biological process "biogenesis" and "signaling", the molecular function "transporting", and the cellular component "organelle" among others (Fig. S4.3). I performed additional searches against a custom database containing 486 accessions of 196 candidate pigmentation genes (hereafter, "pigmentation gene database"), which resulted in significant hits to 5,766 transcripts, and mapping and annotation of 3,088 transcripts in Blast2Go. Of these, 2,400 had not successfully mapped in searches against the nr database (Fig. S4.5). I recovered 184 of the 196 pigmentation genes in the pigmentation gene database. The combined number of mapped transcripts across the two databases was 13,087.

Most of the GO terms recovered from searches against the pigmentation gene database were the same ones recovered from blast searches against the nr database (Fig. S4.4). Only eight GO terms were recovered in the blastx against the nr database that were not also recovered in the pigmentation gene database. Conversely, only one term,

"nutrient reservoir activity," was recovered in searches against the pigmentation gene database that was not also recovered in searches against the nr database.

*Differential expression profiling*

The expression profiles between the species were largely similar. Few isoforms (Fig. 4.1) and genes (Fig. 4.2) showed significant differences in expression at an FDR of 0.05, of which, only 10 isoforms (Fig. 4.3) and 4 genes (Fig. 4.4) showed greater than four-fold differences. The *de novo* transcript, "TRINITY_DN35505_c0_g12," had a highly significant blast hit to *BLOC1S1* (biogenesis of lysosomal organelles complex 1 subunit 1; chromosome 33 in *Gallus*), an autosomal protein-coding gene that has a known pigmentation function in mice. TRINITY_DN43361_c1_g1_i1 appears to blast to *NDUFB7* (NADH:ubiquinone oxidoreductase subunit B7; chromosome 30 in *Gallus*), a highly conserved autosomal gene. Finally, TRINITY_DN49333_c0_g1_i1 blasted to *SPCS1* (signal peptidase complex subunit 1, chromosome 12 in *Gallus*), an autosomal gene involved in signal peptide processing that is associated with the endoplasmic reticulum. Expression profile TRINITY_DN49852_c0_g5, returned no blast hits. Seven other isoforms were discovered to have 4-fold differences between species (Table 4.2).

Table 4.2. Annotations for the differentially expressed genes and isoforms. The gene symbols and GO terms listed are associated with the blast hits against the nr database.

| Transcript | Level | Gene Symbol | Ontology Terms |
|---|---|---|---|
| TRINITY_DN35505_c0_g12_i1 | both | *BLOC1S1* | mitochondrial intermembrane space, mitochondrial matrix, *BLOC-1* complex, axon cytoplasm, aerobic respiration, endosomal transport, peptidyl-lysine acetylation, neuron projection development, anterograde synaptic vesicle transport, anterograde axonal transport, peptidyl-lysine acetylation |
| TRINITY_DN49852_c0_g5_i1 | both | n.a | n.a. |
| TRINITY_DN43361_c1_g1_i1 | both | *NDUFB7* | mitochondrial respiratory chain complex I, NADH dehydrogenase activity, NADH dehydrogenase |
| TRINITY_DN36544_c0_g2 | gene only | *SPCS1* | signal peptide processing, proteolysis, protein targeting to ER |
| TRINITY_DN49333_c0_g1_i1 | isoform only | *MYBPH* | A band, myosin filament, cell adhesion |
| TRINITY_DN52825_c0_g2_i10 | isoform only | *PSAM5* | Antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent, proteolysis, ubiquitin-dependent protein catabolic process, proteasome-mediated ubiquitin-dependent protein catabolic process, proteolysis involved in cellular protein catabolic process |
| TRINITY_DN54486_c2_g3_i5 | isoform only | *HNRNPA2B1* | nucleoplasm, cytoplasm, membrane, extracellular exosome, catalytic step 2 spliceosome, nucleotide binding, mRNA 3'-UTR binding, miRNA binding, single-stranded telomeric DNA binding, pre-mRNA intronic binding, N6-methyladenosine-containing RNA binding, negative regulation of transcription by RNA polymerase II, mRNA splicing, via spliceosome, mRNA export from nucleus, primary miRNA processing, negative regulation of mRNA splicing, via spliceosome, miRNA transport |

Table continued.

| Transcript | Level | Gene Symbol | Ontology Terms |
|---|---|---|---|
| TRINITY_DN47870_c0_g1_i3 | isoform only | *EMC6* | ER membrane protein complex subunit 6 |
| TRINITY_DN53818_c0_g1_i4 | isoform only | *ACHE* | basement membrane, extracellular space, Golgi apparatus, plasma membrane, cell surface, cell junction, anchored component of membrane, neuromuscular junction, perinuclear region of cytoplasm, acetylcholinesterase activity, collagen binding, serine hydrolase activity, acetylcholine binding, protein homodimerization activity, laminin binding, protein self-association, regulation of receptor recycling, osteoblast development, acetylcholine catabolic process, cell adhesion, receptor internalization, neurotransmitter receptor biosynthetic process, protein tetramerization, retina development in camera-type eye |
| TRINITY_DN54279_c0_g1_i2 | isoform only | *SCAP* | Golgi membrane, endoplasmic reticulum membrane, ER to Golgi transport vesicle membrane, integral component of membrane, cholesterol binding, cholesterol metabolic process, SREBP signaling pathway |
| TRINITY_DN49167_c0_g1_i4 | isoform only | *MRPS25* | mitochondrial inner membrane, ribosome, structural constituent of ribosome, mitochondrial translational elongation, mitochondrial translational termination |

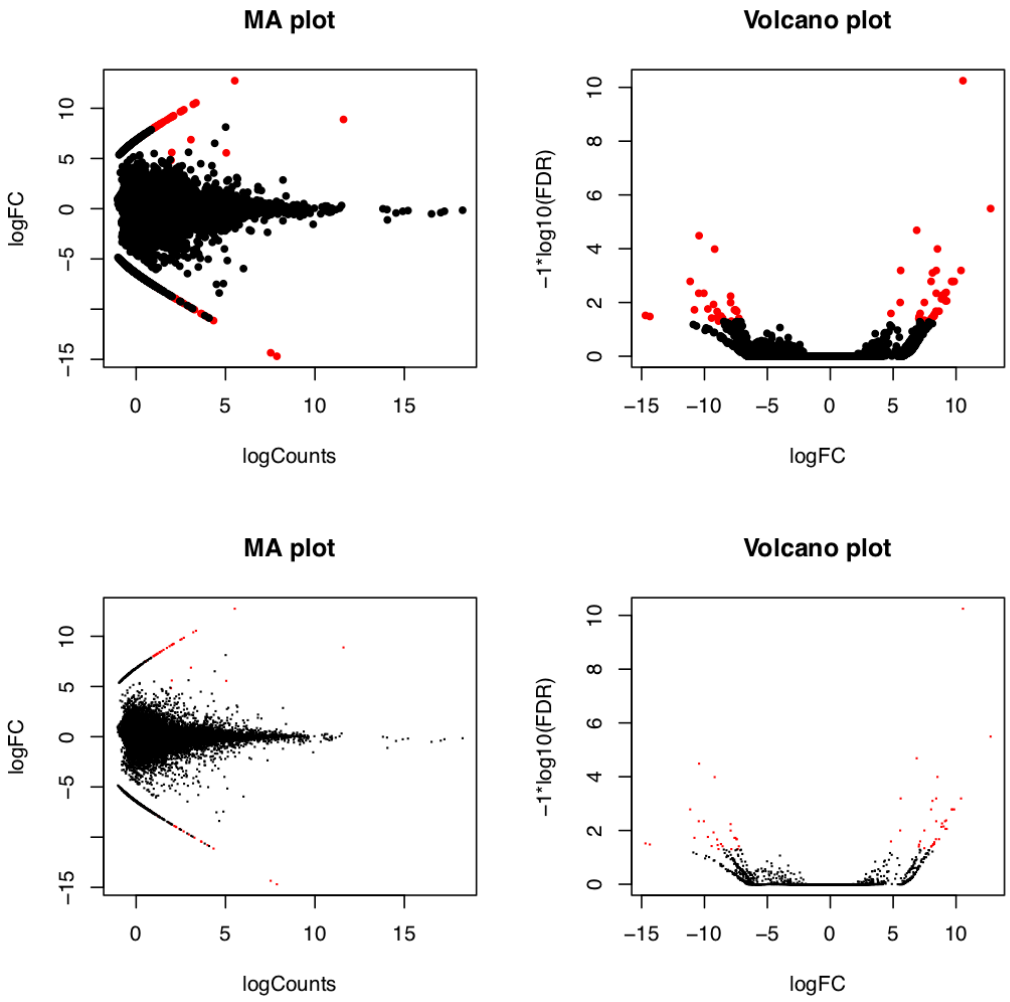



Fig. 4.1. MDS and volcano plot for differentially expressed isoforms between species. DEIs are indicated in red, and are significant with a FDR < 0.05.

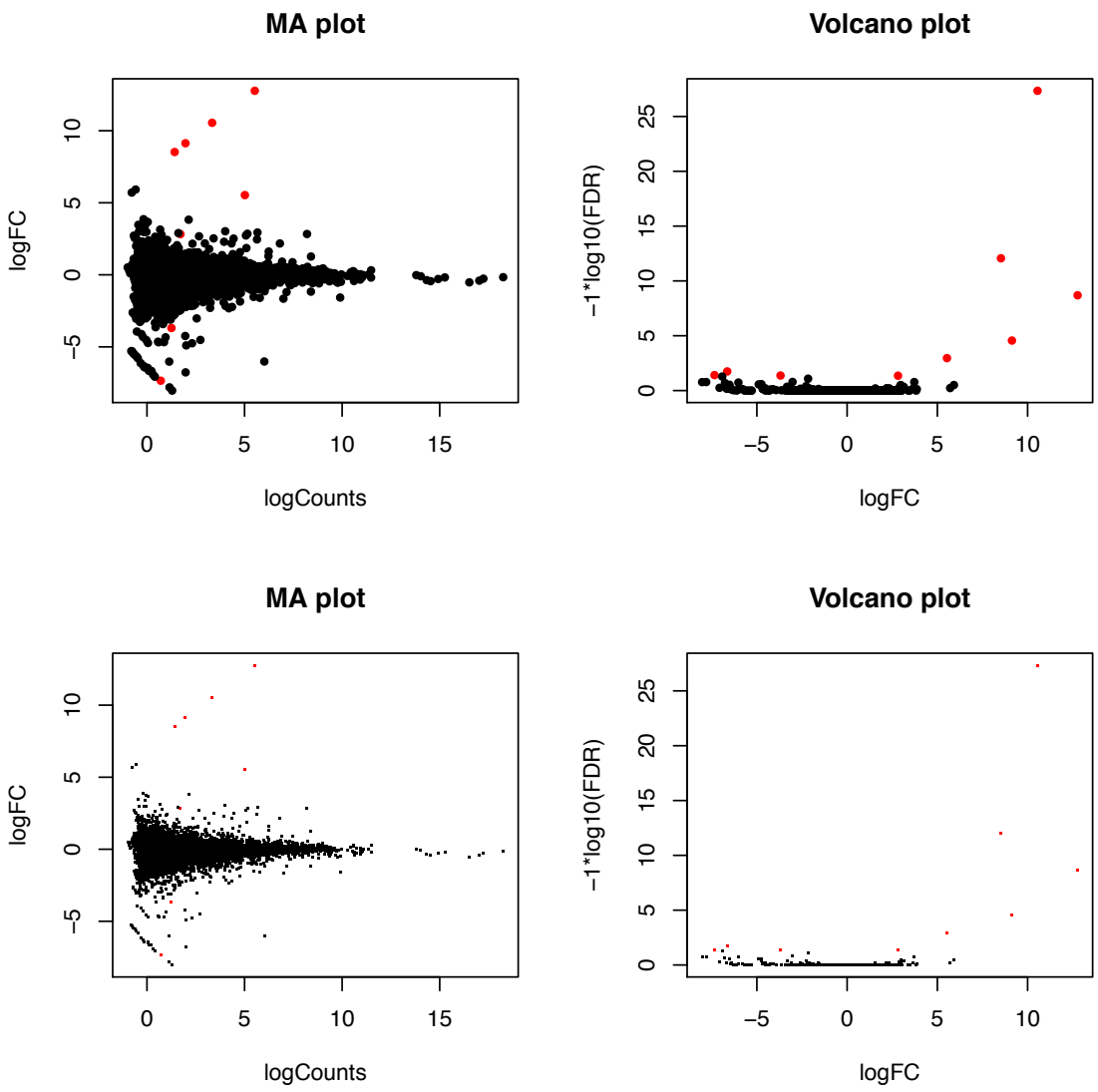



Fig. 4.2. MDS and volcano plot for differentially expressed genes between species. DEGs are indicated in red, and are significant with a FDR < 0.05.

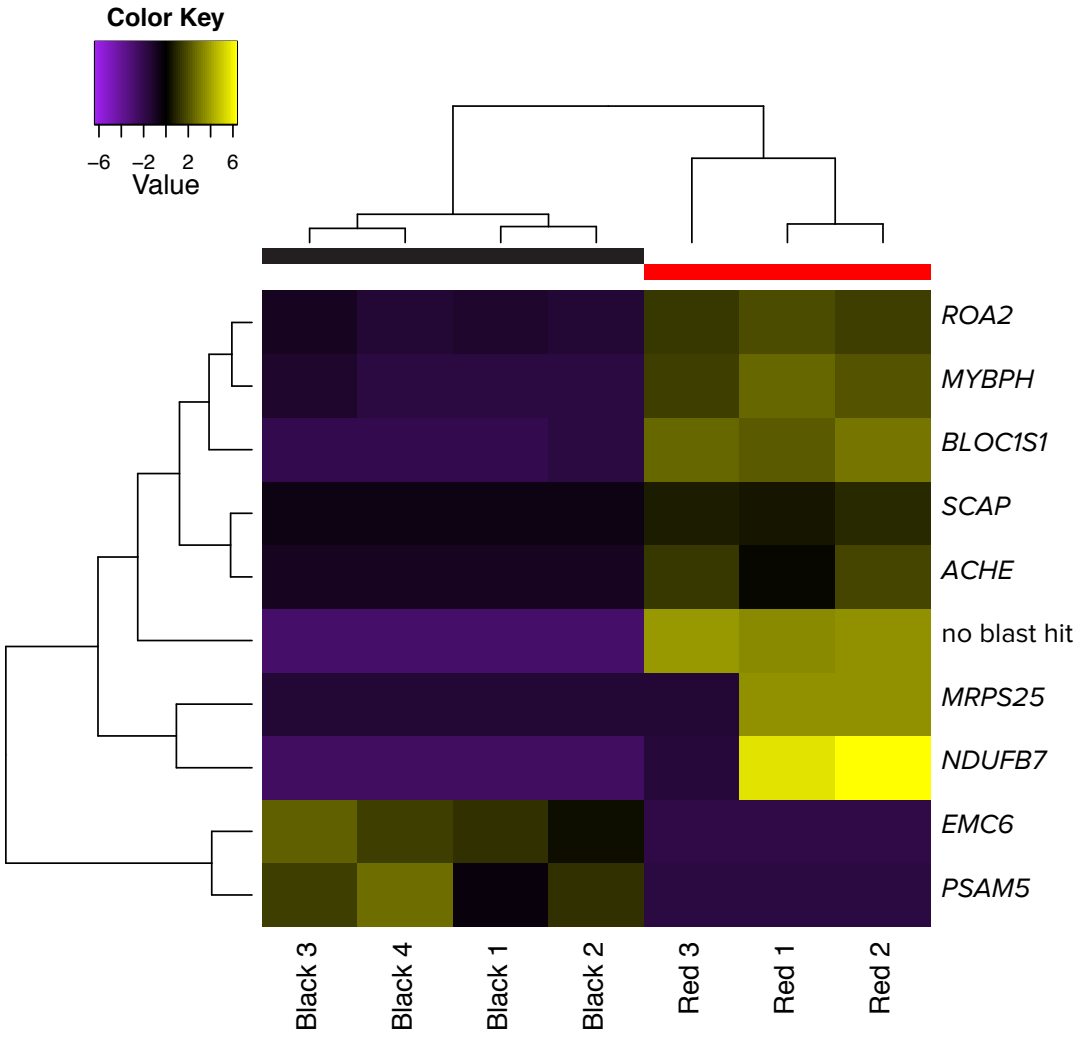



Fig. 4.3. Heat map of 4 differentially expressed isoforms between *Trochilus polytmus* and *T. scitulus*. DEIs are filtered for FDR > 0.05, *p*-values < 0.001 and a log fold change > 4.

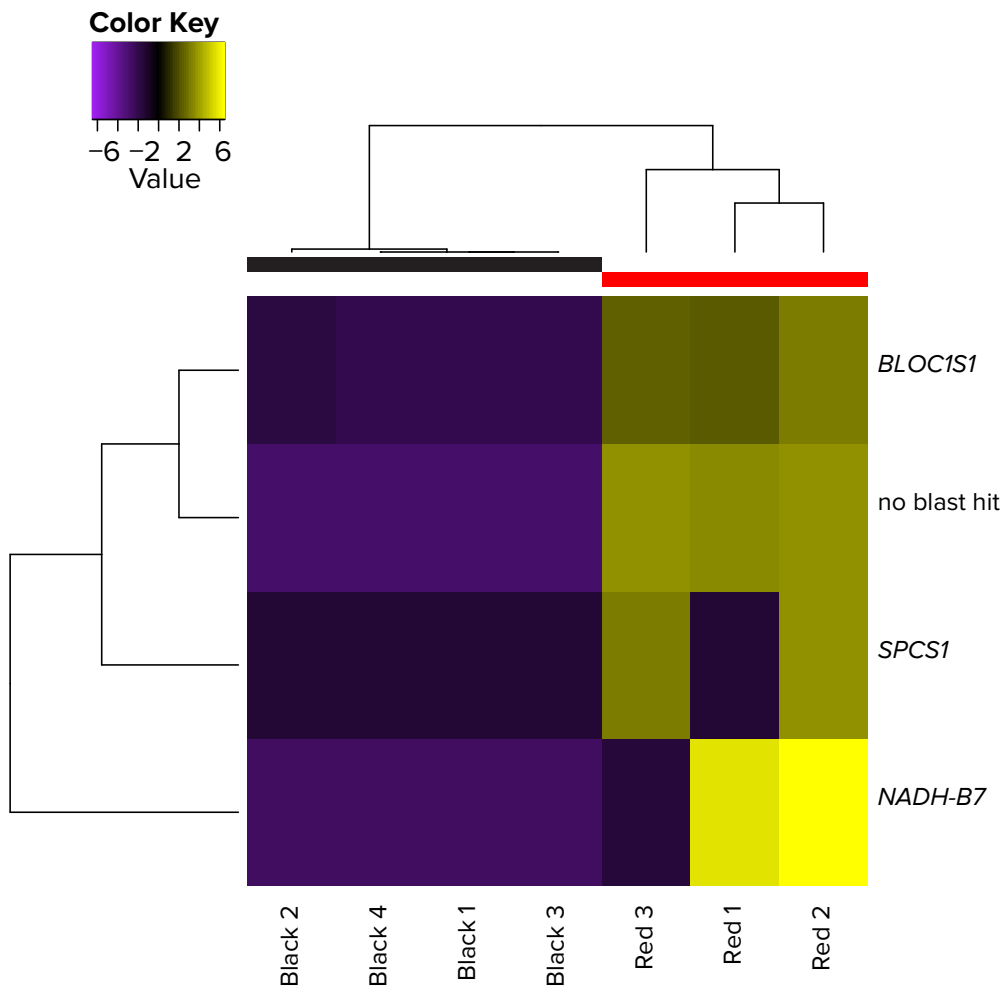



Fig. 4.4. Heat map of differentially expressed genes between *Trochilus polytmus* and *T. scitulus*. DEGs are filtered for FDR > 0.05, *p*-values < 0.001 and a log fold change > 4.

*SNP analysis:*

The PCA of the 155,827 SNPs showed weak but detectable separation in ordinate space between *T. polytmus* and *T. scitulus* along the first PC axis (Fig. 4.5). The percent variation explained for the first PC is 18% and is 17% for the second PC. The *find.clusters* function was not able to correctly assign species labels to all individuals. However, a DAPC using apriori definitions showed high discrimination ability between

species (Fig. 4.6), and 100% or nearly 100% assignments to one or the other genetic cluster (Fig. 4.7).

The Bayescan analysis showed little evidence for positive or diversifying selection; no significant $F_{ST}$ outliers were identified (Fig. 4.5). The $F_{ST}$ coefficient with the highest probability density interval for *T. scitulus* was 0.020 [0.0181, 0.023], and for *T. polytmus* was 0.22 [0.02, 0.025] (Fig. S4.6).

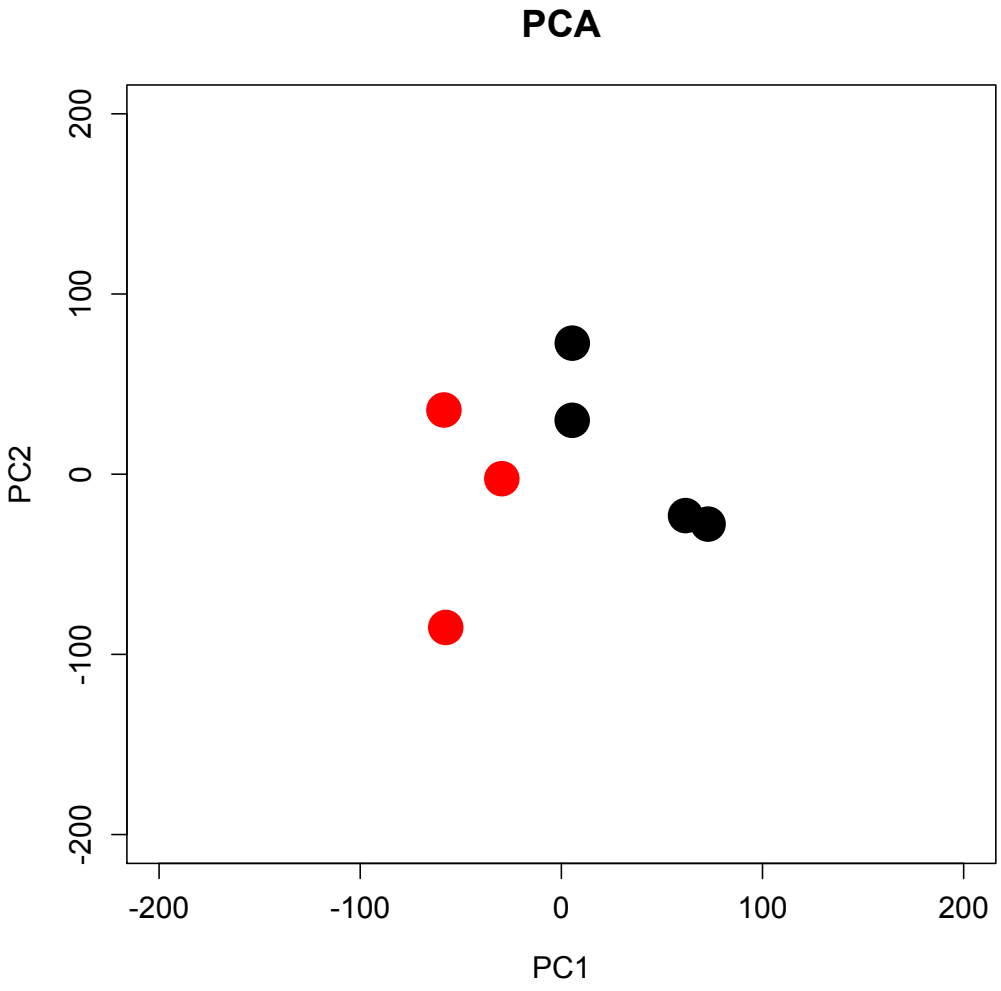



Fig. 4.5. PCA of 155,827 bi-allelic SNPs. *Trochilus polytmus* individuals (n =3) are indicated in red, and *T. scitulus* individuals (n = 4) are indicated in black.

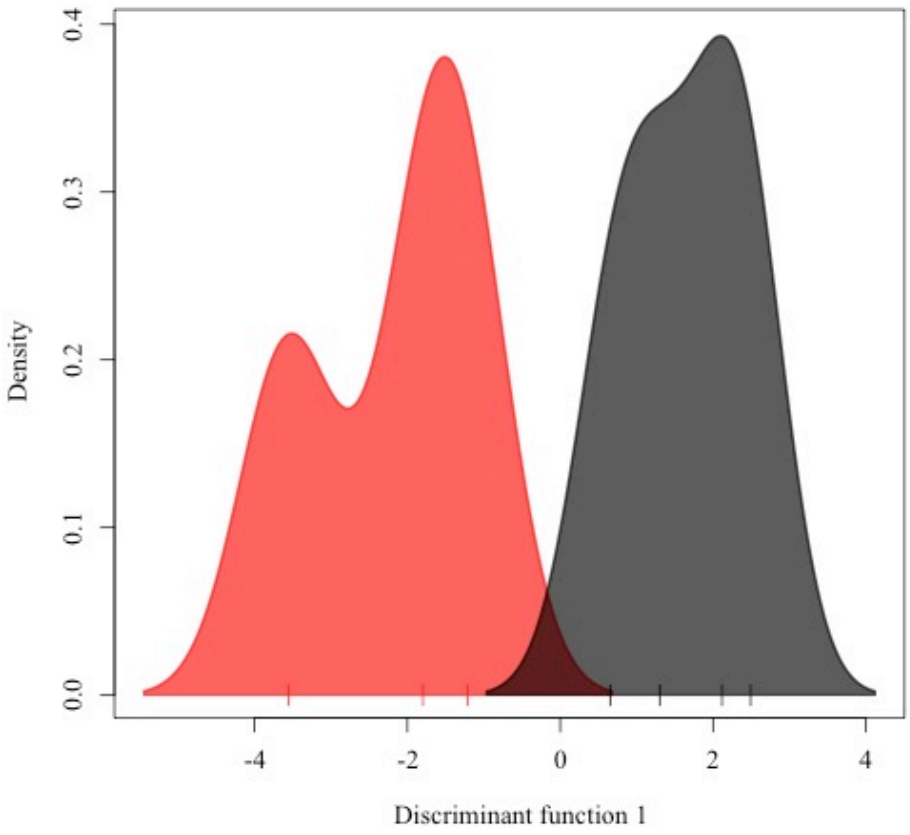



Fig. 4.6. Densities plot of the discriminant function. Red indicates the position of red-billed individuals, and black indicates the position of black-billed individuals along the first discriminant function.

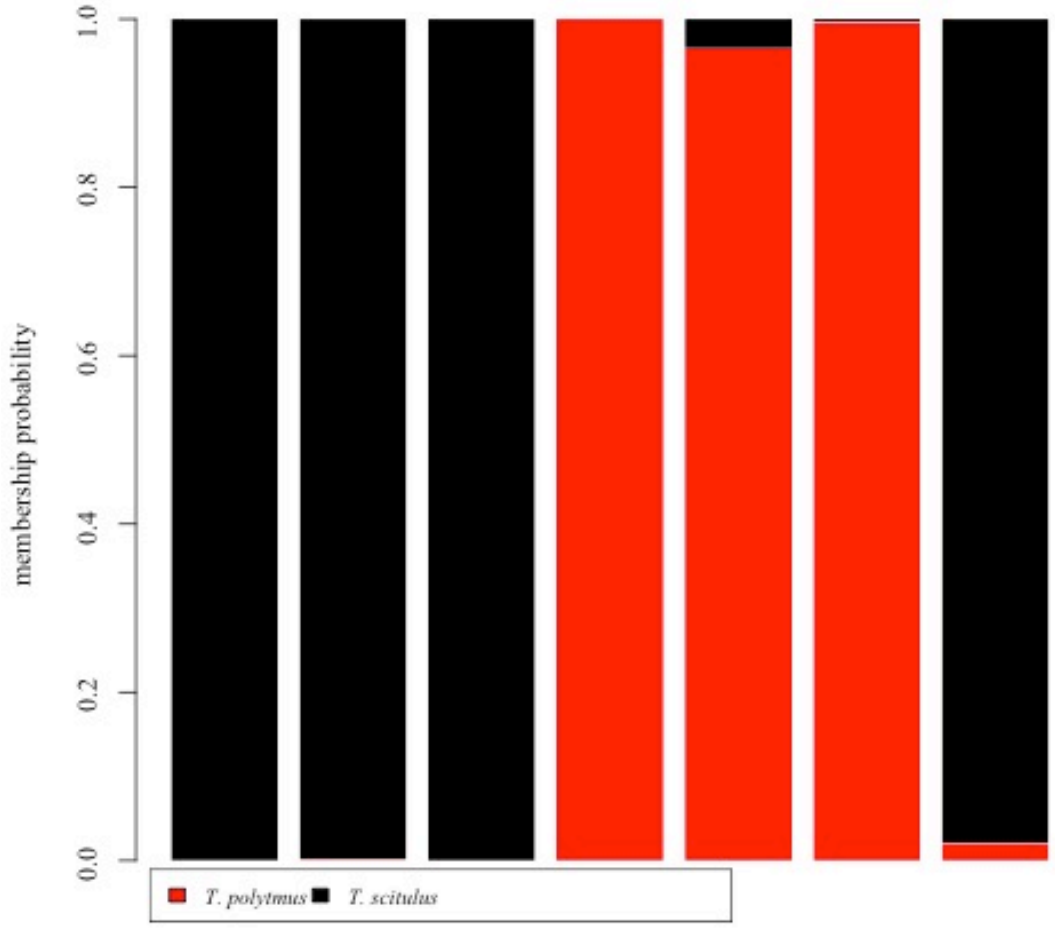


Fig. 4.7. Composition plot reflecting membership probability to the two genetic clusters corresponding to species.

**DISCUSSION**

This study characterized gene expression divergence and nucleotide sequence divergence across the coding genomes of *Trochilus polytmus* and *T. scitulus,* with an eye toward identifying candidate genes that underpin male bill color, a secondary sexual ornament that may be a melanin-based trait. Significant gene expression differences were found in four genes. Intriguingly, one of the differentially expressed genes (*BLOC-1S1*) is part of a gene complex (biogenesis of lysosome-related organelles complex, *BLOC*) that is associated with reduced pigmentation disorders in mice (discussed below). However, the function of *BLOC-1* is not well studied in birds. Nucleotide sequence variation in a

panel of SNPs revealed only modest differentiation between species and little, if any, evidence for positive selection. Below I discuss these results in more detail.

*Transcriptome assembly and annotation*

The total number of mapped genes (11,847) represents approximately two-thirds of the total number of known protein coding genes in other avian reference genomes (e.g. zebra finch, 17,475 protein-coding genes; Warren et al. 2010). The contig statistics including the N50 (814 bp) and ExN50 (1,587 bp) indicate that the overall quality of the transcriptome was good. However, the low number of recovered full-length genes (n = 7,426) and relatively low proportion of proper pairs per library (< 70 % in six of seven libraries) indicate that the assembly is fragmented or that coverage was insufficient to fully sequence coding genes (discussed below).

Fragmentation can result from degradation of RNA in the tissue sample prior to stabilization. Degradation increases with increasing PMI in field-based collection of avian tissues (Cheviron et al. 2011). Interestingly, the one sample (CDJ 238) that was excluded from the study due to poor library quality had the longest PMI of all the samples (10 minutes), and correspondingly, the lowest yield of mRNA (14 ng). Thus, it appears this sample may have been more heavily affected by degradation than the other samples. The loss of this sample is unfortunate but serves as a good benchmark for future ornithological expeditions and field-based investigations. As a general rule of thumb, tissues should be extracted and stabilized in less than 10 minutes, and preferably in less than 5 minutes. Therefore, it is possible to collect RNA-quality tissue while processing a bird as a museum skin, but efficiency of the preparator is important.

The ExN50 statistics indicate that a small number of transcripts were extremely abundant: four transcripts account for over 50% of the total expression. Two of these four mapped to ribosomal RNA (rRNA) molecules, and the other two mapped to mtDNA molecules. Thus, it appears that the libraries contained some rRNA contamination, despite using an mRNA-specific extraction protocol. Contamination by other highly abundant species of RNA, such as rRNA, is a known issue for the mRNA DIRECT protocol (Invitrogen). The final temperature-dependent step is a goldilocks point, with lower temperatures having more specificity to mRNA targets, and higher temperatures resulting in a higher yield of eluted mRNA. While rRNA contamination is undesirable because it takes sequencing reads away from the target mRNA reads, it is also unlikely to bias the downstream differential expression analysis, since the TMM normalization step should compensate for these extreme values (Brian Haas, Broad Institute, pers. comm.). In this study, however, it may have significantly reduced the number of recovered genes. Additional sequencing of the same libraries may prove useful in terms of recovering additional genes.

*Gene expression profiling*

Overall, gene expression profiles were extremely similar between the two species of *Trochilus*, supporting the hypothesis that expression levels of most genes evolve under stabilizing selection (Gilad et al. 2006). Only 4 genes and 10 isoforms were differentially expressed. One of the four differentially expressed genes identified, *BLOC-1S1*, is included in the list of 95 candidate pigmentation genes identified by Poelstra et al. (2013), but is otherwise obscure in the literature for vertebrate pigmentation. For

example, it is not mentioned in major reviews of vertebrate pigmentation (Mundy 2005; Hoekstra 2006; Hubbard et al. 2010; Galván and Solano 2016).

*BLOC-1S1*, a component of the ubiquitously expressed '*BLOC-1'* or biogenesis of lysosome-related organelles complex (Falcon-Perez et al. 2002), is a 200 kDa ubiquitously expressed protein complex containing proteins encoded by *BLOC-1S1* and at least seven other genes: *DTNBP1*, *Muted*, *Pallidin*, *Cappuccino*, *Snapin*, *BLOC-1S2*, and *BLOC-1S3* (Falcon-Perez et al. 2002; Starcevic and Dell'Angelica 2004; Morris et al. 2008). *BLOC-1* is required for normal biogenesis of specialized organelles such as melanosomes and platelet dense granules (Starcevic and Dell'Angelica 2004). The genes that encode these proteins are defective in mouse strains that serve as models of Hermansky-Pudlak Syndrome (HPS), a genetic disorder characterized by hypopigmentation and platelet storage pool deficiency (Starcevic and Dell'Angelica 2004). *BLOC*-complex genes also have associations with schizophrenia (Morris et al. 2008). The normal protein levels for *BLOC-1* are reduced in pallid mice, which have a mutation in *Pallidin* and an associated missense mutation in *BLOC-3* (Starcevic and Dell'Angelica 2004). Additionally, *BLOC-1* has known physical and functional interactions with *AP-3*. Biochemical, genetic, and functional data are consistent with a model in which *BLOC-1* functions with *AP-3* and *BLOC-2* in trafficking of *TYRP1* from endosomes to melanosomes (Di Pietro et al. 2006).

*BLOC-1S1* is upregulated in *T. polytmus* and downregulated in *T. scitulus*. While assuming a causal link between differential expression in this gene and divergence in bill color is purely speculative, the fact that one of four differentially expressed genes is a candidate pigmentation gene is noteworthy. In the *de novo* transcriptome, 12 Trinity

"genes", each with a single isoform, were annotated as BLOC-1S1. Of these only one "gene" was differentially expressed. The high number of trinity genes (versus isoforms) is interesting and may reflect a need for better coverage at this gene to better resolve the sequence variants.

*SNP analysis*

The PCA of the RNAseq panel of 155,827 high-quality SNPs showed only modest separation in ordinate space between *Trochilus polytmus* and *T. scitulus*, indicating low sequence divergence across the coding genomes, on par with levels reported from a previous population genetic analysis of a RADseq dataset (Chapter 3). Such low divergence is likely due to the young age of the lineages and ongoing hybridization, the latter of which is expected to obscure or erase the signal for divergence at neutral loci. Evidence for hybridization in *Trochilus* has been documented by the presence of morphological hybrids (Gill et al. 1973; Graves 2009, 2015) and signatures of genetic admixture in hybrid zone populations (Chapter 3). The between-species component of genetic variation was significant, however, and a DAPC was able to discriminate the species fairly well.

I performed an $F_{ST}$ outlier analysis in Bayescan to identify genetic variants that may be associated with the bill color differences. Because of the diagnostic differences in plumage, and low background levels of genetic differentiation, loci linked to causal variants should be easily detectable as differentiation outliers (Poelstra et al. 2013). However, the estimate of $F_{ST}$ coefficients in Bayescan for both *T. polytmus* ($F_{ST}$ = 0.22) and *T. scitulus* ($F_{ST}$ = 0.20) were low, and Bayescan did not detect any outliers among the 121,092 biallelic SNPs tested. Additional analyses performed under a lower prior odds

setting (100 vs. 1,000) did not return any outliers. The lack of outliers in both analyses could be due to a lack of statistical power. However, the use of a loci-specific FDR to determine significance is more powerful than using a family-wide correction such as Bonferroni corrections in testing multiple comparisons (Foll et al. 2010). More likely, strongly differentiated loci that show clear signatures of selection may not have been captured by this transcriptomics study.

## REFERENCES


Baião, P. C., E. A. Schreiber, and P. G. Parker. 2007. The genetic basis of the plumage polymorphism in red-footed boobies (*Sula sula*): a Melanocortin-1 Receptor (MC1R) Analysis. J. Hered. 98:287–292.

Brewster, W., and O. Bangs. 1901. On an overlooked species of *Aithurus*. Proc. N. Engl. Zoöl. Club 2:47–50.

Campagna, L., I. Gronau, L. F. Silveira, A. Siepel, and I. J. Lovette. 2015. Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Mol. Ecol. 24:4238–4251.

Cheviron, Z. A., M. D. Carling, and R. T. Brumfield. 2011. Effects of postmortem interval and preservation method on RNA isolated from field-preserved avian tissues. The Condor 113:483–489.

Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676.

Coyne, J. A., and T. D. Price. 2000. Little evidence for sympatric speciation in island birds. Evolution 54:2166–2171.

Darwin, C. 1871. The descent of man, and selection in relation to sex. Princeton University Press, Princeton, New Jersey.

DePristo, M. A., E. Banks, R. E. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M., Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

Di Pietro, S. M., J. M. Falcón-Pérez, D. Tenza, S. R. G. Setty, M. S. Marks, G. Raposo, and E. C. Dell'Angelica. 2006. BLOC-1 interacts with BLOC-2 and the AP-3 complex to facilitate protein trafficking on endosomes. Mol. Biol. Cell 17:4027–4038.

Dunn, C. W., M. Howison, and F. Zapata. 2013. Agalma: an automated phylogenomics workflow. BMC Bioinformatics 14:330.

Falcon-Perez, J., M. Starcevic, R. Gautam, and E. C. Dell'Angelica. 2002. BLOC-1, a novel complex containing the pallidin and muted proteins involved in the biogenesis of melanosomes and platelet-dense granules. J Biol Chem277: 28191-28199. J. Biol. Chem. 277:28191–9.

Fischer, M. C., M. Foll, L. Excoffier, and G. Heckel. 2011. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). Mol. Ecol. 20:1450–1462.

Foll, M., M. C. Fischer, G. Heckel, and L. Excoffier. 2010. Estimating population structure from AFLP amplification intensity. Mol. Ecol. 19:4638–4647.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993.

Galván, I., and F. Solano. 2016. Bird integumentary melanins: biosynthesis, forms, function and evolution. Int. J. Mol. Sci. 17.

Gilad, Y., A. Oshlack, and S. A. Rifkin. 2006. Natural selection on gene expression. Trends Genet. 22:456–461.

Gill, F. B., F. J. Stokes, and C. Stokes. 1973. Contact zones and hybridization in the Jamaican hummingbird, *Trochilus polytmus* (L.). The Condor 75:170–176.

Graves, G. R. 2015. A primer on the hybrid zone of Jamaican streamertail hummingbirds (Trochilidae: *Trochilus* ). Proc. Biol. Soc. Wash. 128:111–124.

Graves, G. R. 2009. Ontogeny of bill color in male streamertail hummingbirds (*Trochilus*). J. Caribb. Ornithol. 44–47.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Harrison, Richard G. 1998. Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. Pp. 19–32 *in* Endless forms: species and speciation. Oxford University Press, New York.

Hoekstra, H. E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 97:222–234.

Hubbard, J. K., J. A. C. Uy, M. E. Hauber, H. E. Hoekstra, and R. J. Safran. 2010. Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet. 26:231–239.

Jax, E., M. Wink, and R. H. S. Kraus. 2018. Avian transcriptomics: opportunities and challenges. J. Ornithol. 159:599–629.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically. BMC Genet. 11:94.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Li, B., and C. N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Mason, N. A., and S. A. Taylor. 2015. Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a holarctic songbird. Mol. Ecol. 24:3009–3025.

Mayr. E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press, New York.

McGuire, J. A., C. C. Witt, J. V. Remsen, A. Corl, D. L. Rabosky, D. L. Altshuler, and R. Dudley. 2014. Molecular phylogenetics and the diversification of hummingbirds. Curr. Biol. 24:910–916.

Morris, D. W., K. Murphy, N. Kenny, S. M. Purcell, K. A. McGhee, S. Schwaiger, J.-M. Nangle, G. Donohoe, S. Clarke, P. Scully, J. Quinn, D. Meagher, P. Baldwin, N. Crumlish, E. O'Callaghan, J. L. Waddington, M. Gill, and A. P. Corvin. 2008. Dysbindin (DTNBP1) and the biogenesis of lysosome-related organelles complex 1 (BLOC-1): main and epistatic gene effects are potential contributors to schizophrenia susceptibility. Biol. Psychiatry 63:24–31.

Mundy, N. I. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc. R. Soc. B Biol. Sci. 272:1633–1640.

Nosil, P., and J. L. Feder. 2013. Genome evolution and speciation: toward quantitative descriptions of pattern and process. Evolution 67:2461–2467.

Poelstra, J. W., H. Ellegren, and J. B. W. Wolf. 2013. An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. Heredity 111:467–473.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Muller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414.

Powell, T. H. Q., G. R. Hood, M. O. Murphy, J. S. Heilveil, S. H. Berlocher, P. Nosil, and J. L. Feder. 2013. Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Diptera: Tephritidae). Evolution 67:2561–2576.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Robinson, M. D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–192.

Shaw, K. L., and S. P. Mullen. 2011. Genes versus phenotypes in the study of speciation. Genetica 139:649–661.

Starcevic, M., and E. C. Dell'Angelica. 2004. Identification of snapin and three novel proteins (BLOS1, BLOS2, and BLOS3/reduced pigmentation) as subunits of biogenesis of lysosome-related organelles complex-1 (BLOC-1). J. Biol. Chem. 279:28393–28401.

The Gene Ontology Consortium. 2015. Gene ontology consortium: going forward. Nucleic Acids Res. 43:D1049–D1056.

Theron, E., K. Hawkins, E. Bermingham, R. E. Ricklefs, and N. I. Mundy. 2001. The molecular basis of an avian plumage polymorphism in the wild: A melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. Curr. Biol. 11:550–557.

Toews, D. P. L., L. Campagna, S. A. Taylor, C. N. Balakrishnan, D. T. Baldassarre, P. E. Deane-Coe, M. G. Harvey, D. M. Hooper, D. E. Irwin, C. D. Judy, N. A. Mason, J. E. McCormack, K. G. McCracken, C. H. Oliveros, R. J. Safran, E. S. C. Scordato, K. F. Stryjewski, A. Tigano, J. A. C. Uy, and B. M. Winger. 2015. Genomic approaches to understanding population divergence and speciation in birds. The Auk 133:13–30.

Uebbing, S., A. Künstner, H. Mäkinen, N. Backström, P. Bolivar, R. Burri, L. Dutoit, C. F. Mugal, A. Nater, B. Aken, P. Flicek, F. J. Martin, S. M. J. Searle, and H. Ellegren. 2016. Divergence in gene expression within and between two closely related flycatcher species. Mol. Ecol. 25:2015–2028.

Uy, J. A. C., E. A. Cooper, S. Cutie, M. R. Concannon, J. W. Poelstra, R. G. Moyle, and C. E. Filardi. 2016. Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers. Proc. R. Soc. B Biol. Sci. 283:20160731.

Uy, J. A. C., D. E. Irwin, and M. S. Webster. 2018. Behavioral isolation and incipient speciation in birds. Annu. Rev. Ecol. Evol. Syst., doi: 10.1146/annurev-ecolsys-110617-062646.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y.-C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S.-P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, and R. K. Wilson. 2010. The genome of a songbird. Nature 464:757–762.

Whitehead, A., and D. L. Crawford. 2006. Variation within and among species in gene expression: raw material for evolution. Mol. Ecol. 15:1197–1211.

Wit, P., M. H. Pespeni, J. T. Ladner, D. J. Barshis, F. Seneca, H. Jaris, N. O. Therkildsen, M. Morikawa, and S. R. Palumbi. 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Mol. Ecol. Resour. 12:1058–1067.

Wolf, J. B. W., T. Bayer, B. Haubold, M. Schilhabel, P. Rosenstiel, and D. Tautz. 2010. Nucleotide divergence vs. gene expression differentiation: comparative

transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. Mol. Ecol. 19:162–175.

# CHAPTER 5.

# DISCUSSION

My survey of the black-billed streamertail (*Trochilus scitulus*) confirms that this species is abundant within its range (Chapter 2). Population density has interesting implications for hybrid zone dynamics. If the streamertail hybrid zone is a tension zone maintained by selection-dispersal balance (discussed below), in theory it should be able to move toward geographical regions of low density and dispersal (Barton 1979; Barton and Hewitt 1985). *Trochilus polytmus* are known to be among the most abundant birds on the island of Jamaica (Bond 1936, 1956). If *T. scitulus* had a much lower population density relative to *T. polytmus*, the hybrid zone would be predicted to move eastward, potentially threatening the integrity of *T. polytmus* as a distinct morpho-group. Chapter 2 confirms that the black-billed streamertail population has similarly high densities as its red-billed congener, one factor that might help contribute to the zone's spatial stability over at least 50 generations (Gill et al. 1973; MacColl and Lewis 2000, Judy 2018).

In Chapter 3, I report the first comprehensive study of the morphological and genetic structure of the *Trochilus* hybrid zone. Despite a lack of signal for neutral genomic divergence, I found extremely narrow clines in two morphological traits: bill width and bill color. The cline for male bill color (2.2km) is among the narrowest reported from an avian hybrid zone. While the limited number of divergent traits (morphological and genetic) makes it difficult to diagnose the class of hybrid zone, and thereby infer the type of selection maintaining it, I was able to rule out neutrality as a plausible explanation for cline patterns.

The exceptionally narrow cline for bill color, a secondary sexual character, appears to be under strong selection. Given that it is a sexually selected trait, sexual selection likely plays a prominent role in driving initial speciation in this case.

In Chapter 4, I used a transcriptomics approach to further characterize the genomes and look for genes that may underpin the divergent bill color trait. To do this, I assembled a *de novo* transcriptome and annotated it using blast searches against NCBI's non-redundant protein database and a custom database that I built based on relevant gene ontology terms and searches in the pigmentation gene literature. I performed mapping and annotation using Blast2Go. I used alignments of individual libraries to the *de novo* transcriptome to make comparisons between species regarding significant differences in the expression of each gene. I found that the two species are similar in their expression profiles, which likely reflects the young age of the lineages. The lack of expression divergence could also signal stabilizing selection, but I cannot readily assess this hypothesis under such short evolutionary timescales. However, strong differences (greater than four-fold log differences) were significantly different in four genes.

One of these genes, *BLOC-1S1*, is a known pigmentation gene. Proteins associated with this and other genes in the *BLOC*-complex are ubiquitously expressed across tissue types. Mutations and abnormal function of the *BLOC*-complex are associated with hypopigmentation in laboratory strains of inbred mice that are models for Hermansky-Pudlak syndrome (Di Pietro et al. 2006). The role *of BLOC-1S1* in avian pigmentation is not well studied; however, it is mentioned in an exhaustive candidate gene search that supported studies of plumage polymorphism between hooded and

carrion crows (Poelstra et al. 2014). Its potential role in affecting expression of bill pigmentation in *Trochilus* warrants further investigation.

I also identified SNPs from the raw RNAseq reads using the GATK pipeline. A principal components analysis of the entire SNP dataset revealed modest but detectable divergence between the species, similar to the Radseq dataset in Chapter 3. A Bayesian $F_{ST}$ outlier analysis detected no outliers and returned little evidence for positive selection. Thus, regions of strong genomic divergence are few (or absent) and may be in regulatory regions that were not captured by this study.

In sum, I characterized the morphological and genetic patterns of divergence between incipient species of streamertail hummingbirds using population genetic techniques and cline models. I was also able to sequence the coding genomes of these incipient species, to get a handle on the genetic basis of divergence and look for candidate genes to explain bill color, which appears to be a melanin-based trait. The results from the genetic datasets (microsatellites, RadSeq SNPs, RNA-seq SNPs) confirm low background levels of genomic divergence between these species. However, the hybrid zone between them appears to be stable, and the two morphological traits are concordant and narrow relative to neutral expectations. All cline models coincide with the Rio Grande Valley, which may represent a dispersal barrier for these birds. While distinguishing between primary and secondary intergradation is challenging at this shallow time scale, it seems likely that the zone is maintained in selection-dispersal balance, as are the majority of vertebrate hybrid zones. If so, strong selection against hybrid male bill color, a sexual ornament, may be what primarily maintains the zone.

This study provides an important genomic snapshot of early-stage speciation in a non-model bird.

## REFERENCES


Barton, N. H. 1979. The dynamics of hybrid zones. Heredity 43:341–359.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Bond, J. 1936. Birds of the West Indies. Academy of Natural Sciences of Philadelphia, Philadelphia.

Bond, J. 1956. Check-list of the birds of the West Indies. Academy of Natural Sciences of Philadelphia, Philadelphia.

Di Pietro, S. M., J. M. Falcón-Pérez, D. Tenza, S. R. G. Setty, M. S. Marks, G. Raposo, and E. C. Dell'Angelica. 2006. BLOC-1 interacts with BLOC-2 and the AP-3 complex to facilitate protein trafficking on endosomes. Mol. Biol. Cell 17:4027–4038.

Gill, F. B., F. J. Stokes, and C. Stokes. 1973. Contact zones and hybridization in the Jamaican hummingbird, *Trochilus polytmus* (L.). The Condor 75:170–176.

Judy, C. D. in press. Density and abundance in the black-billed streamertail (*Trochilus scitulus*) in eastern Jamaica. Journal Caribbean Ornith.

MacColl, A., and S. Lewis. 2000. Hybridisation and ecology of Jamaican streamertail hummingbirds. BirdLife Jam. 4–10.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Muller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414.

**APPENDIX A.**

**SUPPLEMENTARY MATERIAL FOR CHAPTER 3**

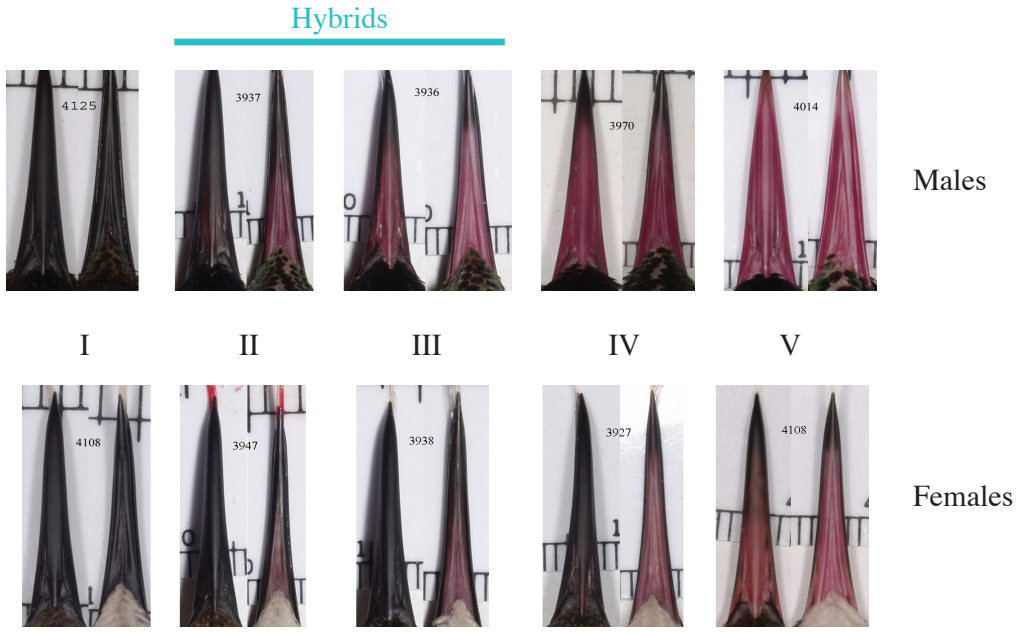



Fig. S3.1. Hybrid index (Graves 2015) to classify adult *T. scitulus* (I), hybrid (II, III), and *T. polytmus* (IV,V) individuals. Dorsal (left) and ventral (right) views in each pair of images.

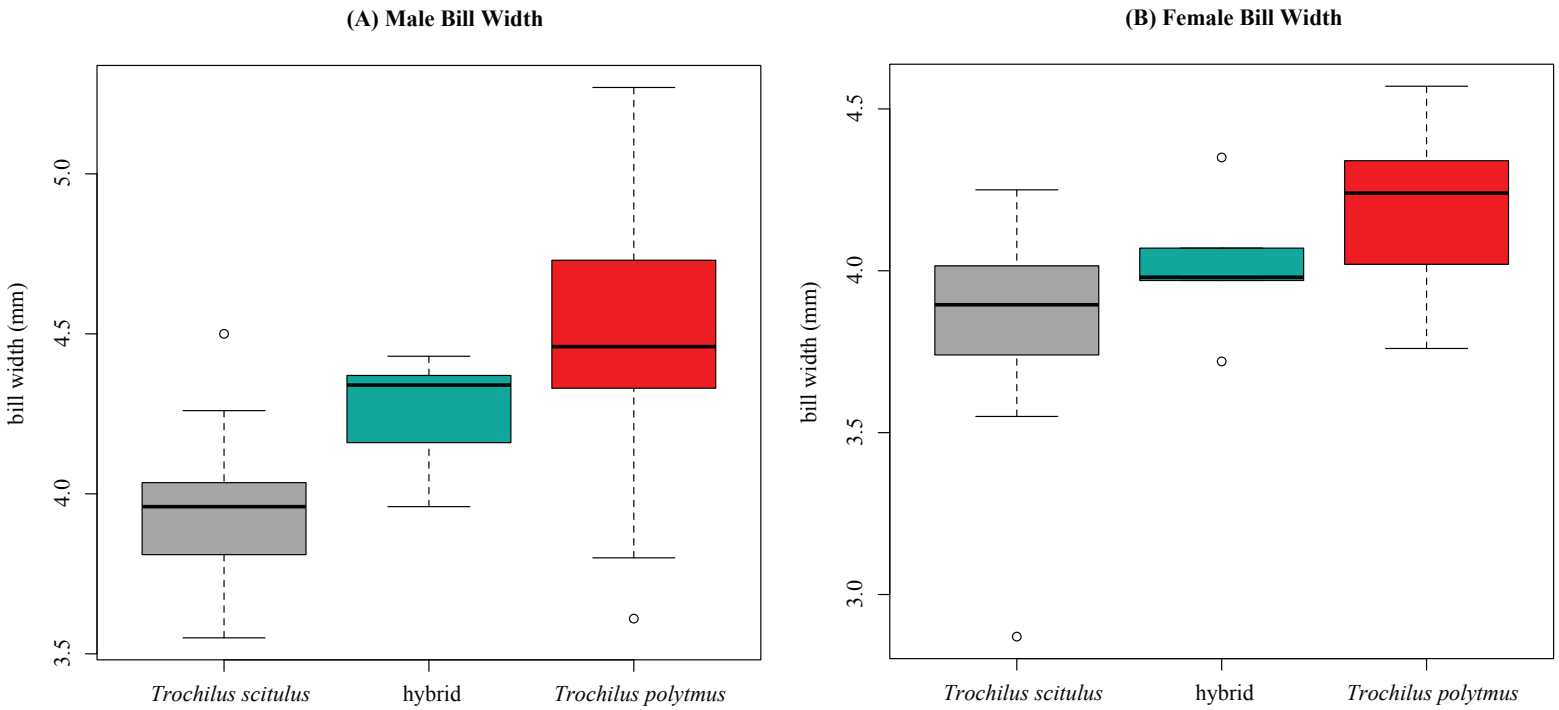



Fig. S3.2. Male (A) and female (B) bill width in *T. scitulus*, hybrid, and *T. polytmus* individuals.

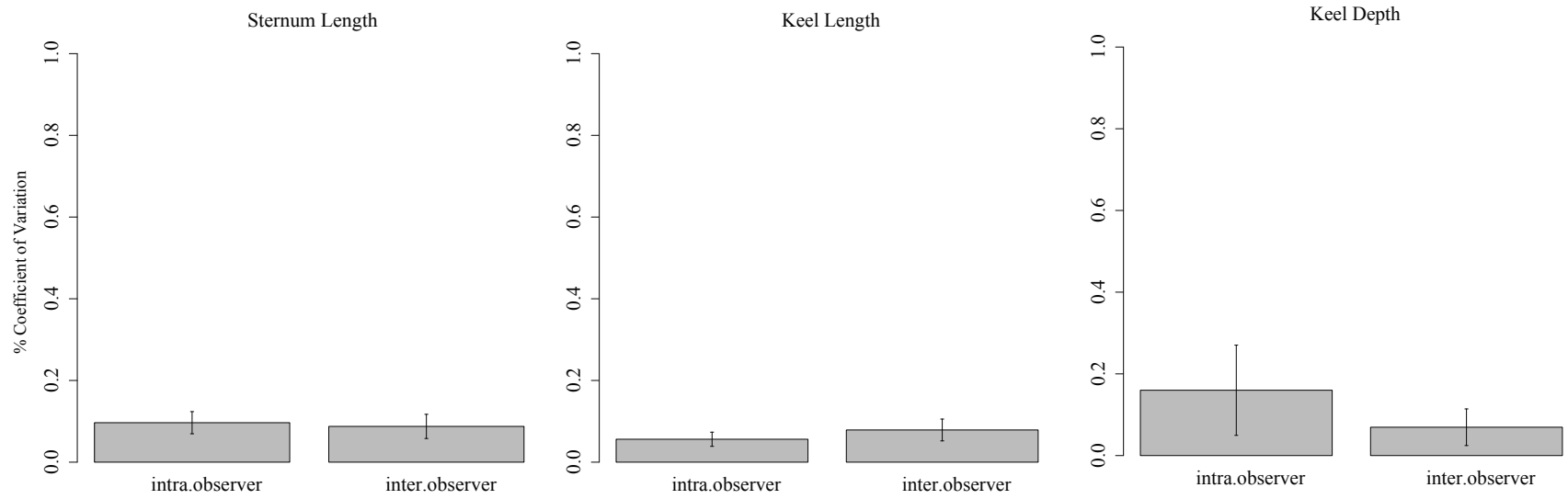



Fig. S3.3. Coefficient of variation within and between observers in three skeletal elements: sternum length, keel length, and keel depth. Substantial observer bias was not detected in these characters.

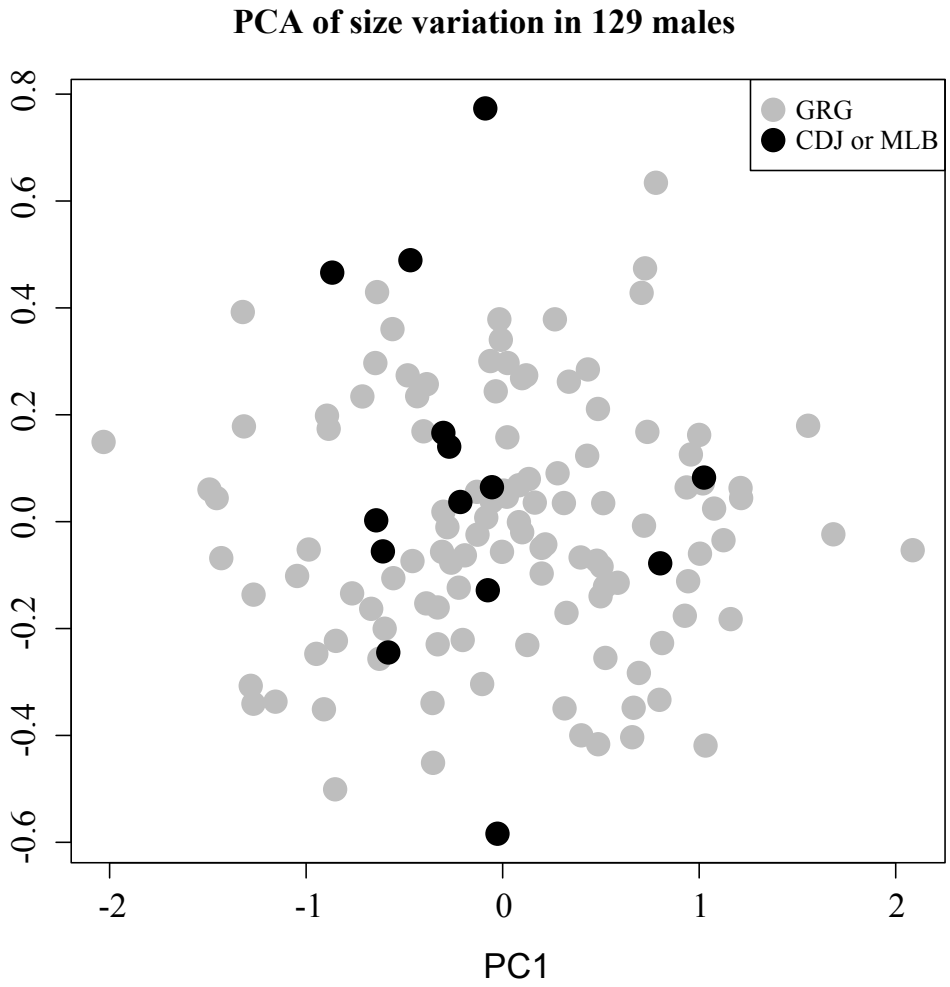



Fig. S3.4. PCA of size variation in 129 males for seven measured skeletal elements (three of the seven elements were used in the current study: sternum length, keel length, and keel depth). I measured the individuals labeled with black points, which were collected by CDJ or MLB, and GRG measured the individuals labeled with gray points that he collected. No artificial grouping was detected in the PCA.

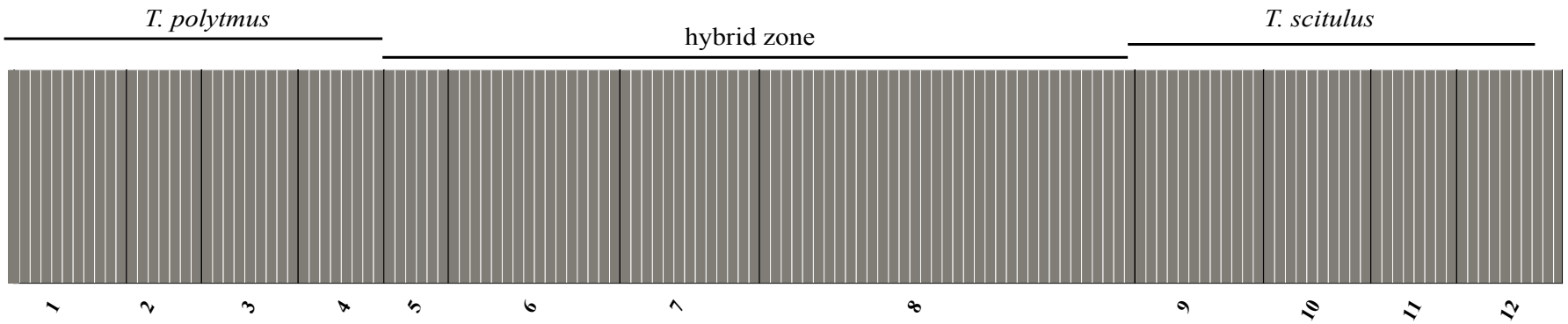



Fig. S3.5. Admixture proportions assigned by FASTSTRUCTURE based on K = 2 for 6,451 SNPs in 145 individuals of *T. polytmus*, *T. scitulus*, and their hybrids in 12 sites. All individuals showed nearly 100% assignment to a single genetic cluster.

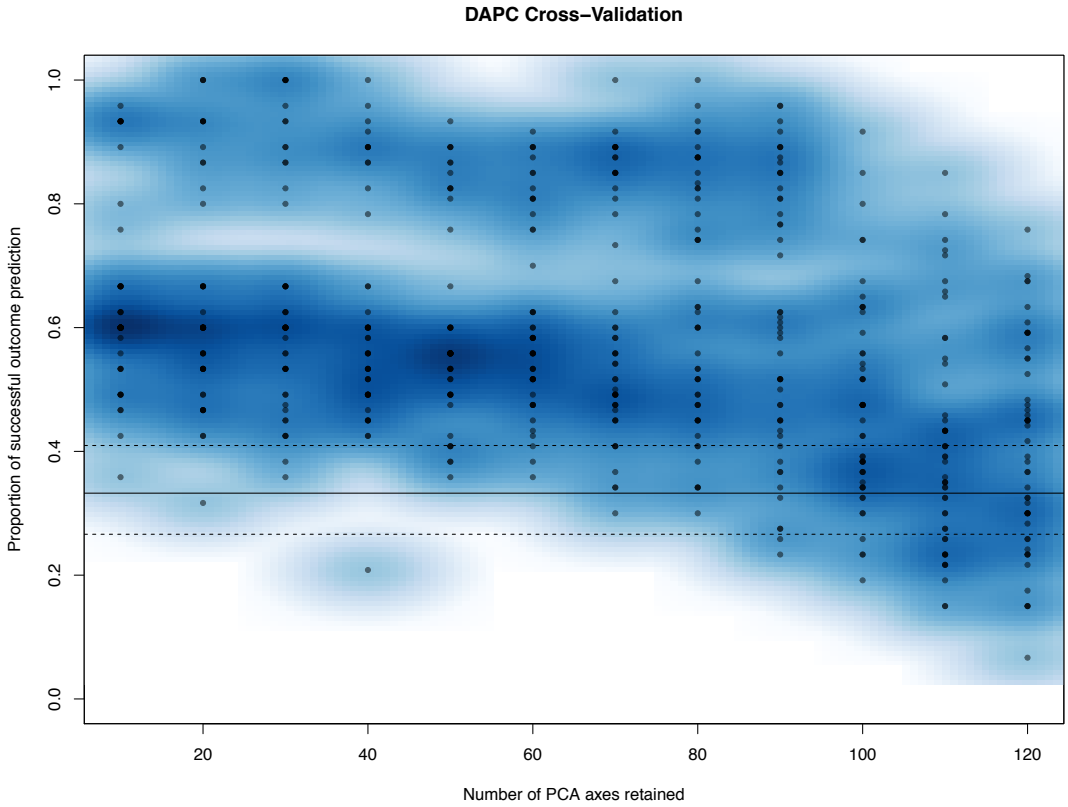



Fig. S3.6. Cross-validation optimization procedure results for 30 repetitions. The lowest RMSE is associated with 30 retained PCs.

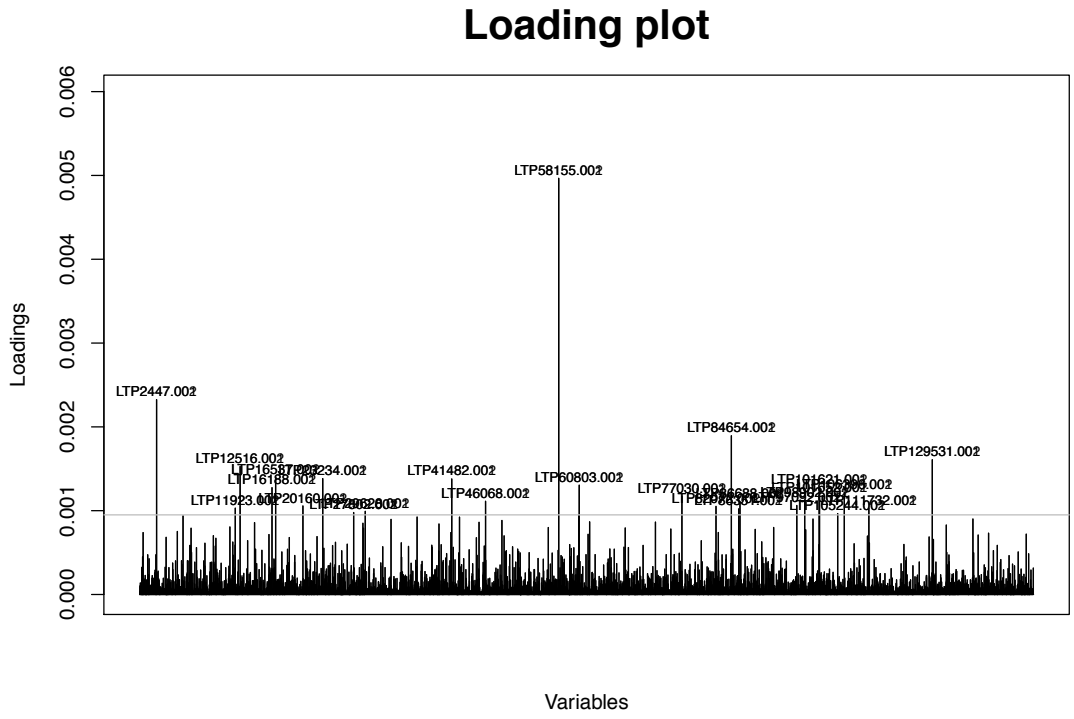



Fig. S3.7. Loading plot of allelic contributions of 6,451 SNPs to the DAPC. Alleles are lined up horizontally along the x-axis, with each bar in the plot representing a unique allele. There are 46 alleles (23 loci) that have contributions greater than 0.001, the threshold indicated by the horizontal gray line.

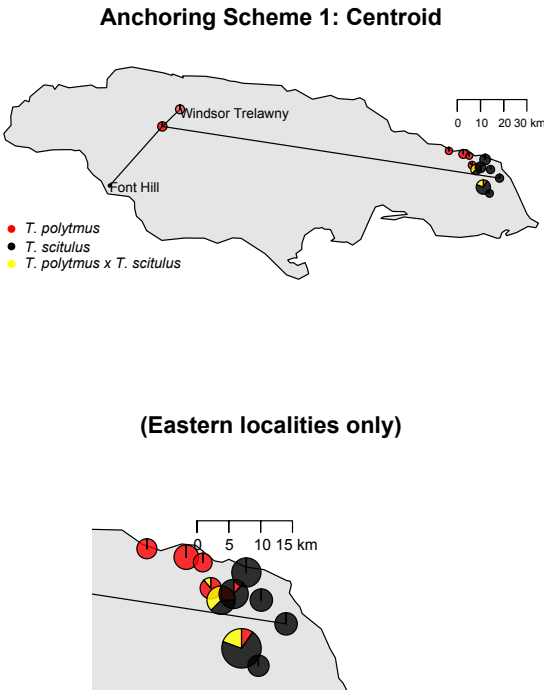
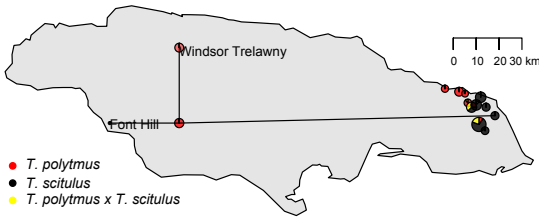



Fig. S3.8 A. The 'Centroid' anchor is the average latitudinal and longitudinal coordinate among all sampled individuals grouped into Site 1. These individuals are pooled from two localities: Windsor Trelawney (n=7) and Font Hill (n=3).

**Anchoring Scheme 2: Right Angle**




**(Eastern localities only)**

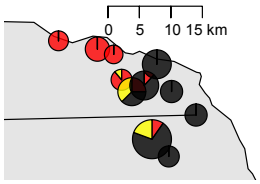



Fig. S3.8 B. The 'Right Angle' anchor is a geographic locality that forms a right angle between Site 1 localities of Windsor Trelawney (n=7) and Font Hill (n=3).

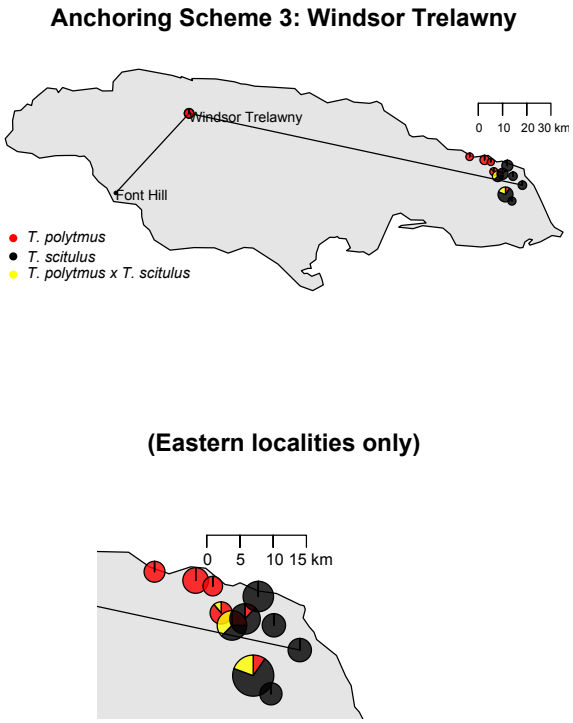



Fig. S3.8 C. The ‘Windsor Trelawny’ anchor is on the Site 1 locality Windsor Trelawney where the majority of Site 1 individuals were collected.

Table S3.1. Specimen information for all specimens used in this study. Preparators: Gary R. Graves = GRG, Caroline D. Judy = CDJ, and Matthew L. Brady = MLB. North, West = geographical coordinates of sites. Species: *Trochilus polytmus* = 1, *Trochilus scitulus* = 2, and phenotypic hybrids = 3. Sex: Female = F, Male = M. Bill striations = N.A. for specimens where poor photo quality rendered bill striations difficult to evaluate with certainty. All other columns: Y = Specimen was used in the specified analysis.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ref. *T. polytmus* | 4117 | NMNH | 635693 | GRG | 18.058 | -77.940 | Dec-05 | 1 | F | | Y | Y | Y | Y | Y |
| 1 | ref. *T. polytmus* | 4119 | NMNH | 635695 | GRG | 18.062 | -77.940 | Dec-05 | 1 | F | | Y | Y | Y | Y | Y |
| 1 | ref. *T. polytmus* | 4159 | NMNH | 636171 | GRG | 18.359 | -77.659 | Mar-06 | 1 | F | | Y | Y | Y | Y | |
| 1 | *ref. T. polytmus* | CVD_236 | LSUMNS | 29498 | CDJ | 18.354 | -77.647 | Dec-14 | 1 | F | | Y | Y | | | |
| 1 | ref. *T. polytmus* | CVD_237 | LSUMNS | 29499 | CDJ | 18.354 | -77.647 | Dec-14 | 1 | F | Y | | | | | |
| 1 | *ref. T. polytmus* | CVD_240 | IOJ | 29505 | CDJ | 18.358 | -77.643 | Dec-14 | 1 | F | NA | | | | | |
| 1 | ref. *T. polytmus* | CVD_241 | LSUMNS | 29507 | CDJ | 18.354 | -77.647 | Dec-14 | 1 | F | | Y | Y | | | |
| 1 | *ref. T. polytmus* | MLB_363 | LSUMNS | 29500 | MLB | 18.354 | -77.647 | Dec-14 | 1 | F | | Y | Y | Y | | |
| 1 | ref. *T. polytmus* | MLB_364 | LSUMNS | 29501 | MLB | 18.358 | -77.643 | Dec-14 | 1 | F | Y | | | | | |
| 1 | *ref. T. polytmus* | 4118 | NMNH | 635694 | GRG | 18.062 | -77.940 | Dec-05 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | ref. *T. polytmus* | 4157 | NMNH | 636169 | GRG | 18.364 | -77.666 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | *ref. T. polytmus* | 4158 | NMNH | 636170 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ref. *T. polytmus* | 4160 | NMNH | 636172 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | *ref. T. polytmus* | 4161 | NMNH | 636173 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | ref. *T. polytmus* | 4162 | NMNH | 636174 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | *ref. T. polytmus* | 4163 | NMNH | 636175 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | ref. *T. polytmus* | 4164 | NMNH | 636176 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | |
| 1 | *ref. T. polytmus* | 4165 | NMNH | 636177 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | ref. *T. polytmus* | 4166 | NMNH | 636178 | GRG | 18.359 | -77.660 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 1 | *ref. T. polytmus* | CVD_238 | LSUMNS | 29502 | CDJ | 18.358 | 77.643 | Dec-14 | 1 | M | | | | | | |
| 1 | ref. *T. polytmus* | CVD_239 | LSUMNS | 29504 | CDJ | 18.354 | 77.647 | Dec-14 | 1 | M | | | | | | |
| 1 | *ref. T. polytmus* | CVD_242 | LSUMNS | 29510 | CDJ | 18.354 | -77.647 | Dec-14 | 1 | M | NA | | | | | |
| 1 | ref. *T. polytmus* | MLB_365 | IOJ | 29503 | MLB | 18.354 | -77.647 | Dec-14 | 1 | M | NA | | | | | |
| 1 | *ref. T. polytmus* | MLB_366 | LSUMNS | 29506 | MLB | 18.354 | -77.647 | Dec-14 | 1 | M | | Y | Y | Y | | |
| 1 | ref. *T. polytmus* | MLB_367 | LSUMNS | 29508 | MLB | 18.354 | -77.647 | Dec-14 | 1 | M | | | | | | |
| 2 | *ref. T. polytmus* | 4008 | NMNH | 633622 | GRG | 18.195 | -76.554 | Feb-04 | 1 | F | | Y | Y | Y | Y | Y |
| 2 | ref. *T. polytmus* | 4009 | NMNH | 633623 | GRG | 18.195 | -76.555 | Feb-04 | 1 | F | | Y | Y | Y | Y | |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | *ref. T. polytmus* | 4013 | NMNH | 633627 | GRG | 18.195 | -76.554 | Feb-04 | 1 | F | | Y | Y | Y | Y | Y |
| 2 | ref. *T. polytmus* | 4007 | NMNH | 633621 | GRG | 18.195 | -76.554 | Feb-04 | 1 | M | | Y | Y | Y | Y | Y |
| 2 | *ref. T. polytmus* | 4010 | NMNH | 633624 | GRG | 18.196 | -76.553 | Feb-04 | 1 | M | | Y | Y | Y | Y | Y |
| 2 | ref. *T. polytmus* | 4011 | NMNH | 633625 | GRG | 18.196 | -76.554 | Feb-04 | 1 | M | | Y | Y | Y | Y | Y |
| 2 | *ref. T. polytmus* | 4012 | NMNH | 633626 | GRG | 18.196 | -76.553 | Feb-04 | 1 | M | | Y | Y | Y | Y | Y |
| 2 | ref. *T. polytmus* | 4014 | NMNH | 633628 | GRG | 18.196 | -76.555 | Feb-04 | 1 | M | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4005 | NMNH | 633619 | GRG | 18.178 | -76.499 | Feb-04 | 1 | F | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4006 | NMNH | 633620 | GRG | 18.179 | -76.497 | Feb-04 | 1 | F | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4130 | NMNH | 636142 | GRG | 18.185 | -76.496 | Mar-06 | 1 | M | | Y | Y | Y | Y | |
| 3 | *T. polytmus* | 4131 | NMNH | 636143 | GRG | 18.184 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4132 | NMNH | 636144 | GRG | 18.184 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4133 | NMNH | 636145 | GRG | 18.185 | -76.496 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4134 | NMNH | 636146 | GRG | 18.185 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4135 | NMNH | 636147 | GRG | 18.185 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4136 | NMNH | 636148 | GRG | 18.184 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | *T. polytmus* | 4137 | NMNH | 636149 | GRG | 18.184 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | |
| 3 | *T. polytmus* | 4138 | NMNH | 636150 | GRG | 18.185 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 3 | *T. polytmus* | 4139 | NMNH | 636151 | GRG | 18.185 | -76.495 | Mar-06 | 1 | M | | Y | Y | Y | Y | |
| 4 | *T. polytmus* | 4140 | NMNH | 636152 | GRG | 18.176 | -76.471 | Mar-06 | 3 | M | Y | | | | Y | Y |
| 4 | *T. polytmus* | 4141 | NMNH | 636153 | GRG | 18.176 | -76.472 | Mar-06 | 3 | M | Y | | | | Y | Y |
| 4 | *T. polytmus* | 4142 | NMNH | 636154 | GRG | 18.176 | -76.472 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 4 | *T. polytmus* | 4143 | NMNH | 636155 | GRG | 18.176 | -76.472 | Mar-06 | 3 | M | Y | | | | Y | Y |
| 4 | *T. polytmus* | 4144 | NMNH | 636156 | GRG | 18.176 | -76.472 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 4 | *T. polytmus* | 4145 | NMNH | 636157 | GRG | 18.176 | -76.471 | Mar-06 | 1 | M | | Y | Y | Y | Y | |
| 4 | *T. polytmus* | 4146 | NMNH | 636158 | GRG | 18.176 | -76.471 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 4 | *T. polytmus* | 4147 | NMNH | 636159 | GRG | 18.176 | -76.471 | Mar-06 | 1 | M | | Y | Y | Y | Y | |
| 4 | *T. polytmus* | 4148 | NMNH | 636160 | GRG | 18.176 | -76.471 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 4 | *T. polytmus* | 4149 | NMNH | 636161 | GRG | 18.176 | -76.471 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 5 | hybrid zone | 4105 | NMNH | 635681 | GRG | 18.140 | -76.460 | Dec-05 | 3 | F | | Y | Y | | Y | Y |
| 5 | hybrid zone | 4106 | NMNH | 635682 | GRG | 18.139 | -76.459 | Dec-05 | 1 | F | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | hybrid zone | 4108 | NMNH | 635684 | GRG | 18.140 | -76.459 | Dec-05 | 1 | F | | Y | Y | Y | Y | Y |
| 5 | hybrid zone | 4109 | NMNH | 635685 | GRG | 18.140 | -76.459 | Dec-05 | 1 | F | | Y | Y | Y | Y | Y |
| 5 | hybrid zone | 4102 | NMNH | 635678 | GRG | 18.140 | -76.460 | Dec-05 | 1 | M | | Y | Y | Y | Y | |
| 5 | hybrid zone | 4103 | NMNH | 635679 | GRG | 18.140 | -76.460 | Dec-05 | 1 | M | | Y | Y | Y | Y | |
| 5 | hybrid zone | 4104 | NMNH | 635680 | GRG | 18.140 | -76.460 | Dec-05 | 1 | M | | Y | Y | Y | Y | Y |
| 5 | hybrid zone | 4107 | NMNH | 635683 | GRG | 18.139 | -76.459 | Dec-05 | 1 | M | | Y | Y | Y | Y | |
| 5 | hybrid zone | 4110 | NMNH | 635686 | GRG | 18.140 | -76.459 | Dec-05 | 1 | M | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 3935 | NMNH | 633549 | GRG | 18.110 | -76.447 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 3938 | NMNH | 633552 | GRG | 18.110 | -76.448 | Nov-03 | 3 | F | | Y | Y | | Y | Y |
| 6 | hybrid zone | 3939 | NMNH | 633553 | GRG | 18.110 | -76.447 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 3971 | NMNH | 633585 | GRG | 18.131 | -76.442 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 3972 | NMNH | 633586 | GRG | 18.131 | -76.443 | Feb-04 | 1 | F | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 3973 | NMNH | 633587 | GRG | 18.131 | -76.442 | Feb-04 | 3 | F | | Y | Y | | Y | Y |
| 6 | hybrid zone | 3974 | NMNH | 633588 | GRG | 18.131 | -76.442 | Feb-04 | 1 | F | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 4095 | NMNH | 635671 | GRG | 18.131 | -76.442 | Dec-05 | 2 | F | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | hybrid zone | 4096 | NMNH | 635672 | GRG | 18.131 | -76.443 | Dec-05 | 2 | F | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 4097 | NMNH | 635673 | GRG | 18.131 | -76.442 | Dec-05 | 2 | F | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | CVD_251 | LSUMNS | 29525 | CDJ | 18.137 | 76.438 | Dec-14 | 3 | F | Y | | | | | |
| 6 | hybrid zone | MLB_375 | LSUMNS | 29526 | MLB | 18.137 | 76.438 | Dec-14 | 2 | F | | | | | | |
| 6 | hybrid zone | 3936 | NMNH | 633550 | GRG | 18.110 | -76.447 | Nov-03 | 3 | M | | Y | Y | | Y | Y |
| 6 | hybrid zone | 3937 | NMNH | 633551 | GRG | 18.110 | -76.447 | Nov-03 | 3 | M | | Y | Y | | Y | Y |
| 6 | hybrid zone | 3941 | NMNH | 633555 | GRG | 18.110 | -76.447 | Nov-03 | 3 | M | | Y | Y | | Y | Y |
| 6 | hybrid zone | 3970 | NMNH | 633584 | GRG | 18.131 | -76.442 | Feb-04 | 1 | M | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 4098 | NMNH | 635674 | GRG | 18.131 | -76.443 | Dec-05 | 1 | M | | Y | Y | Y | Y | Y |
| 6 | hybrid zone | 4099 | NMNH | 635675 | GRG | 18.131 | -76.443 | Dec-05 | 3 | M | | Y | Y | | Y | Y |
| 6 | hybrid zone | CVD_250 | LSUMNS | 29522 | CDJ | 18.137 | -76.438 | Dec-14 | 3 | M | | | | | | |
| 6 | hybrid zone | CVD_252 | LSUMNS | 29527 | CDJ | 18.137 | -76.438 | Dec-14 | 1 | M | | Y | Y | Y | | |
| 6 | hybrid zone | MLB_373 | LSUMNS | 29523 | MLB | 18.137 | -76.438 | Dec-14 | 2 | M | | | | | | |
| 6 | hybrid zone | MLB_374 | LSUMNS | 20524 | MLB | 18.137 | -76.438 | Dec-14 | 3 | M | | | | | | |
| 7 | hybrid zone | 3975 | NMNH | 633589 | GRG | 18.137 | -76.429 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | hybrid zone | 3976 | NMNH | 633590 | GRG | 18.137 | -76.429 | Feb-04 | 2 | F | | Y | Y | Y | Y | |
| 7 | hybrid zone | 3981 | NMNH | 633595 | GRG | 18.124 | -76.421 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 3982 | NMNH | 633596 | GRG | 18.123 | -76.421 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 3983 | NMNH | 633597 | GRG | 18.123 | -76.421 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 3984 | NMNH | 633598 | GRG | 18.123 | -76.421 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 3985 | NMNH | 633599 | GRG | 18.126 | -76.420 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 3977 | NMNH | 633591 | GRG | 18.136 | -76.429 | Feb-04 | 2 | M | | Y | Y | Y | Y | |
| 7 | hybrid zone | 3978 | NMNH | 633592 | GRG | 18.137 | -76.429 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 3979 | NMNH | 633593 | GRG | 18.136 | -76.429 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 3980 | NMNH | 633594 | GRG | 18.136 | -76.429 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 4100 | NMNH | 635676 | GRG | 18.137 | -76.429 | Dec-05 | 2 | M | | Y | Y | Y | Y | |
| 7 | hybrid zone | 4101 | NMNH | 635677 | GRG | 18.137 | -76.429 | Dec-05 | 2 | M | | Y | Y | Y | Y | |
| 7 | hybrid zone | 4153 | NMNH | 636165 | GRG | 18.126 | -76.420 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 4154 | NMNH | 636166 | GRG | 18.125 | -76.421 | Mar-06 | 1 | M | | Y | Y | Y | Y | Y |
| 7 | hybrid zone | 4155 | NMNH | 636167 | GRG | 18.136 | -76.429 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | hybrid zone | 4156 | NMNH | 636168 | GRG | 18.137 | -76.429 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3925 | NMNH | 633539 | GRG | 18.062 | -76.418 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3927 | NMNH | 633541 | GRG | 18.062 | -76.419 | Nov-03 | 1 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3929 | IOJ | NA | GRG | 18.062 | -76.418 | Nov-03 | 2 | F | | | | | Y | Y |
| 8 | hybrid zone | 3932 | NMNH | 633546 | GRG | 18.046 | -76.403 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3934 | NMNH | 633548 | GRG | 18.044 | -76.404 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3944 | NMNH | 633558 | GRG | 18.064 | -76.426 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3945 | NMNH | 633559 | GRG | 18.064 | -76.426 | Nov-03 | 3 | F | | Y | Y | | Y | Y |
| 8 | hybrid zone | 3947 | NMNH | 633561 | GRG | 18.064 | -76.426 | Nov-03 | 3 | F | | Y | Y | | Y | Y |
| 8 | hybrid zone | 3953 | NMNH | 633567 | GRG | 18.068 | -76.423 | Nov-03 | 1 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3955 | NMNH | 633569 | GRG | 18.050 | -76.409 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3957 | NMNH | 633571 | GRG | 18.050 | -76.409 | Nov-03 | 2 | F | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3924 | NMNH | 633538 | GRG | 18.062 | -76.419 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3926 | NMNH | 633540 | GRG | 18.045 | -76.403 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3928 | NMNH | 633542 | GRG | 18.062 | -76.418 | Nov-03 | 2 | M | Y | | | | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | hybrid zone | 3930 | NMNH | 633544 | GRG | 18.044 | -76.404 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3931 | NMNH | 633545 | GRG | 18.045 | -76.403 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3933 | NMNH | 633547 | GRG | 18.046 | -76.403 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3940 | NMNH | 633554 | GRG | 18.046 | -76.403 | Nov-03 | 2 | M | Y | | | | Y | Y |
| 8 | hybrid zone | 3942 | NMNH | 633556 | GRG | 18.045 | -76.403 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3943 | NMNH | 633557 | GRG | 18.046 | -76.403 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3946 | NMNH | 633560 | GRG | 18.064 | -76.426 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3948 | NMNH | 633562 | GRG | 18.064 | -76.426 | Nov-03 | 2 | M | Y | | | | Y | Y |
| 8 | hybrid zone | 3949 | NMNH | 633563 | GRG | 18.063 | -76.425 | Nov-03 | 2 | M | | Y | Y | Y | Y | |
| 8 | hybrid zone | 3950 | NMNH | 633564 | GRG | 18.063 | -76.425 | Nov-03 | 1 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3951 | NMNH | 633565 | GRG | 18.068 | -76.432 | Nov-03 | 3 | M | | Y | Y | | Y | |
| 8 | hybrid zone | 3952 | NMNH | 633566 | GRG | 18.068 | -76.423 | Nov-03 | 3 | M | | Y | Y | | Y | Y |
| 8 | hybrid zone | 3954 | NMNH | 633568 | GRG | 18.069 | -76.432 | Nov-03 | 3 | M | | Y | Y | | Y | Y |
| 8 | hybrid zone | 3956 | NMNH | 633570 | GRG | 18.050 | -76.409 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3958 | NMNH | 633572 | GRG | 18.050 | -76.407 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | hybrid zone | 3959 | NMNH | 633573 | GRG | 18.050 | -76.409 | Nov-03 | 2 | M | Y | | | | Y | Y |
| 8 | hybrid zone | 3960 | NMNH | 633574 | GRG | 18.050 | -76.409 | Nov-03 | 2 | M | Y | | | | Y | Y |
| 8 | hybrid zone | 3961 | NMNH | 633575 | GRG | 18.050 | -76.407 | Nov-03 | 2 | M | Y | | | | Y | Y |
| 8 | hybrid zone | 3962 | NMNH | 633576 | GRG | 18.050 | -76.409 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3963 | NMNH | 633577 | GRG | 18.050 | -76.407 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3964 | NMNH | 633578 | GRG | 18.050 | -76.409 | Nov-03 | 2 | M | | Y | Y | Y | Y | |
| 8 | hybrid zone | 3965 | NMNH | 633579 | GRG | 18.050 | -76.407 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3966 | NMNH | 633580 | GRG | 18.050 | -76.407 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 8 | hybrid zone | 3967 | NMNH | 633581 | GRG | 18.050 | -76.409 | Nov-03 | 3 | M | | Y | Y | | Y | Y |
| 8 | hybrid zone | 3968 | NMNH | 633582 | GRG | 18.050 | -76.409 | Nov-03 | 2 | M | Y | | | | Y | |
| 9 | *T. scitulus* | 4000 | NMNH | 633614 | GRG | 18.169 | -76.396 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 4002 | NMNH | 633616 | GRG | 18.169 | -76.396 | Feb-04 | 2 | F | | Y | Y | Y | Y | |
| 9 | *T. scitulus* | 4112 | NMNH | 635688 | GRG | 18.154 | -76.418 | Dec-05 | 2 | F | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 4113 | NMNH | 635689 | GRG | 18.154 | -76.418 | Dec-05 | 2 | F | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 3996 | NMNH | 633610 | GRG | 18.168 | -76.396 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | *T. scitulus* | 3997 | NMNH | 633611 | GRG | 18.168 | -76.396 | Feb-04 | 2 | M | | Y | Y | Y | Y | |
| 9 | *T. scitulus* | 3998 | NMNH | 633612 | GRG | 18.168 | -76.396 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 3999 | NMNH | 633613 | GRG | 18.169 | -76.396 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 4001 | NMNH | 633615 | GRG | 18.169 | -76.396 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 4003 | NMNH | 633617 | GRG | 18.169 | -76.396 | Feb-04 | 2 | M | | Y | Y | Y | Y | |
| 9 | *T. scitulus* | 4004 | NMNH | 633618 | GRG | 18.169 | -76.396 | Feb-04 | 2 | M | | Y | Y | Y | Y | |
| 9 | *T. scitulus* | 4111 | NMNH | 635687 | GRG | 18.154 | -76.419 | Dec-05 | 2 | M | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 4114 | NMNH | 635690 | GRG | 18.155 | -76.419 | Dec-05 | 2 | M | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 4115 | NMNH | 635691 | GRG | 18.155 | -76.418 | Dec-05 | 2 | M | | Y | Y | Y | Y | |
| 9 | *T. scitulus* | 4116 | NMNH | 635692 | GRG | 18.154 | -76.418 | Dec-05 | 2 | M | | Y | Y | Y | Y | Y |
| 9 | *T. scitulus* | 4150 | NMNH | 636162 | GRG | 18.154 | -76.418 | Mar-06 | 2 | M | Y | | | | Y | Y |
| 9 | *T. scitulus* | 4151 | NMNH | 636163 | GRG | 18.155 | -76.418 | Mar-06 | 2 | M | | Y | Y | Y | Y | |
| 9 | *T. scitulus* | 4152 | NMNH | 636164 | GRG | 18.154 | -76.419 | Mar-06 | 2 | M | | Y | Y | Y | | Y |
| 10 | *ref. T. scitulus* | 3914 | NMNH | 633528 | GRG | 18.029 | -76.389 | Oct-03 | 2 | F | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3915 | NMNH | 633529 | GRG | 18.029 | -76.389 | Oct-03 | 2 | F | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | *ref. T. scitulus* | 3918 | NMNH | 633532 | GRG | 18.034 | -76.390 | Oct-03 | 2 | F | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3916 | NMNH | 633530 | GRG | 18.030 | -76.388 | Oct-03 | 2 | M | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3917 | NMNH | 633531 | GRG | 18.034 | -76.390 | Oct-03 | 2 | M | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3919 | NMNH | 633533 | GRG | 18.030 | -76.388 | Oct-03 | 2 | M | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3920 | NMNH | 633534 | GRG | 18.030 | -76.388 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3921 | NMNH | 633535 | GRG | 18.029 | -76.389 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3922 | NMNH | 633536 | GRG | 18.030 | -76.388 | Nov-03 | 2 | M | | Y | Y | Y | Y | Y |
| 10 | *ref. T. scitulus* | 3923 | NMNH | 633537 | GRG | 18.029 | -76.389 | Nov-03 | 2 | M | Y | | | | Y | Y |
| 11 | *ref. T. scitulus* | 3987 | NMNH | 633601 | GRG | 18.123 | -76.384 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 11 | *ref. T. scitulus* | 3989 | NMNH | 633603 | GRG | 18.124 | -76.385 | Feb-04 | 2 | F | | Y | Y | Y | Y | |
| 11 | *ref. T. scitulus* | 3990 | NMNH | 633604 | GRG | 18.121 | -76.384 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 11 | *ref. T. scitulus* | 3991 | NMNH | 633605 | GRG | 18.124 | -76.385 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 11 | *ref. T. scitulus* | 3992 | NMNH | 633606 | GRG | 18.122 | -76.384 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 11 | *ref. T. scitulus* | 3993 | NMNH | 633607 | GRG | 18.124 | -76.385 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |
| 11 | *ref. T. scitulus* | 3994 | NMNH | 633608 | GRG | 18.124 | -76.385 | Feb-04 | 2 | F | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | *ref. T. scitulus* | 3986 | NMNH | 633600 | GRG | 18.121 | -76.384 | Feb-04 | 2 | M | | Y | Y | Y | Y | |
| 11 | *ref. T. scitulus* | 3988 | NMNH | 633602 | GRG | 18.124 | -76.385 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 11 | *ref. T. scitulus* | 3995 | NMNH | 633609 | GRG | 18.123 | -76.384 | Feb-04 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4120 | NMNH | 636132 | GRG | 18.089 | -76.348 | Mar-06 | 2 | F | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4124 | NMNH | 636136 | GRG | 18.089 | -76.348 | Mar-06 | 2 | F | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | CVD_244 | LSUMNS | 29512 | CDJ | 18.108 | 76.333 | Dec-14 | 2 | F | | Y | Y | Y | | |
| 12 | *ref. T. scitulus* | CVD_246 | IOJ | 29517 | CDJ | 18.108 | 76.333 | Dec-14 | 2 | F | NA | | | | | |
| 12 | *ref. T. scitulus* | MLB_370 | LSUMNS | 29515 | MLB | 18.108 | 76.333 | Dec-14 | 2 | F | | | | | | |
| 12 | *ref. T. scitulus* | MLB_371 | LSUMNS | 29516 | MLB | 18.108 | 76.333 | Dec-14 | 2 | F | | | | | | |
| 12 | *ref. T. scitulus* | MLB_372 | LSUMNS | 29520 | MLB | 18.108 | 76.333 | Dec-14 | 1 | F | | Y | Y | Y | | |
| 12 | *ref. T. scitulus* | 4121 | NMNH | 636133 | GRG | 18.090 | -76.347 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4122 | NMNH | 636134 | GRG | 18.090 | -76.347 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4123 | NMNH | 636135 | GRG | 18.089 | -76.348 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4125 | NMNH | 636137 | GRG | 18.090 | -76.347 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4126 | NMNH | 636138 | GRG | 18.089 | -76.348 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |

Table continued.

| Site | Locality Type | Field ID | Museum | Museum ID | Preparator | North | West | Date | Species | Sex | Bill Striations | Bill Color | Bill Width | Body Size | Microsats | GBS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | *ref. T. scitulus* | 4127 | NMNH | 636139 | GRG | 18.090 | -76.347 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4128 | NMNH | 636140 | GRG | 18.090 | -76.347 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | 4129 | NMNH | 636141 | GRG | 18.090 | -76.347 | Mar-06 | 2 | M | | Y | Y | Y | Y | Y |
| 12 | *ref. T. scitulus* | CVD_245 | LSUMNS | 29514 | CDJ | 18.108 | 76.333 | Dec-14 | 2 | M | | | | | | |
| 12 | *ref. T. scitulus* | CVD_247 | LSUMNS | 29518 | CDJ | 18.108 | 76.333 | Dec-14 | 2 | M | | | | | | |
| 12 | *ref. T. scitulus* | CVD_248 | LSUMNS | 29519 | CDJ | 18.108 | 76.333 | Dec-14 | 2 | M | | Y | Y | | | |
| 12 | *ref. T. scitulus* | CVD_249 | LSUMNS | 29521 | CDJ | 18.108 | 76.333 | Dec-14 | 2 | M | | Y | Y | Y | | |
| 12 | *ref. T. scitulus* | MLB_368 | IOJ | 29513 | MLB | 18.108 | 76.333 | Dec-14 | 2 | M | NA | | | | | |
| 12 | *ref. T. scitulus* | MLB_369 | LSUMNS | 29509 | MLB | 18.108 | 76.333 | Dec-14 | 1 | M | | Y | Y | Y | | |

Table S3.2. HZAR parameter estimates for three different anchoring schemes. The best choice model is indicated next to the name of each anchor scheme.

| Parameter | "Windsor" (model I) | "Right" (model IV) | "Centroid" (model I) |
|---|---|---|---|
| center | 132.2 (131.6 - 132.7) | 130.8 (130.6 - 130.9) | 138.4 (137.9 - 138.8) |
| width | 3.33 (2.51 - 4.40) | 0.45 (0.17 - 0.60) | 3.00 (2.19 - 3.94) |
| varH | 4.71 (2.38 - 11.4) | 4.58 (1.97 - 15.0) | 4.64 (2.24 - 12.78) |
| deltaL | n.a | 7.89 E-5 (2.30 E-5 - 0.08) | n.a |
| deltaR | n.a | n.a. | n.a |
| tauL | n.a | 0.73 (0.02 - 0.15) | n.a |
| tauR | n.a | n.a. | n.a |

Table S3.3. Tests of interspecific and intraspecific differences in bill characters. (A) Chi-squared tests for interspecific differences in bill color based on the hybrid index. (B) Analysis of variance for bill width. (C) Post-hoc Tukey Test comparisons to determine which are significant at the 0.05 (*), 0.01 (**), and 0.001 level (***). Comparisons are between 1) male *T. scitulus* and male *T. polytmus*, 2) male and female *T. scitulus*, 3) female *T. scitulus* and female *T. polytmus*, and 4) male and female *T. polytmus*

| A. | $X^2$ Test Statistic | Degrees of Freedom | *p*-value | Significance |
|---|---|---|---|---|
| Interspecific Male | 35 | 2 | 2.51 E -8 | *** |
| Interspecific Female | 22 | 2 | 1.67 E -5 | *** |

| B. | Degrees of Freedom | Sum Squares | Mean Squares | F | *p*-value | Significance |
|---|---|---|---|---|---|---|
| Group | 3 | 9.742 | 3.247 | 59.78 | < 2 E -16 | *** |
| Residuals | 52 | 2.825 | 0.054 | | | |

| C. | Difference | Lower | Upper | Adj. *p*-value | Significance |
|---|---|---|---|---|---|
| 1) Interspecific Male | 0.907 | 0.697 | 1.117 | 0.00 | *** |
| 2) Intraspecific *T. polytmus* | 0.518 | 0.260 | 0.775 | 1.26 E -5 | *** |
| 3) Interspecific Female | 0.504 | 0.231 | 0.777 | 5.60 E -5 | *** |
| 4) Intraspecific *T. scitulus* | 0.114 | -0.114 | 0.342 | 0.55 | NS |

Table S3.4. Tests of interspecific and intraspecific differences in body size. *Trochilus scitulus* = black, and *T. polytmus* = red. (A) Analysis of variance for bill width. (B) Post-hoc Tukey Test comparisons to determine which comparisons are significant at the 0.05 (*), 0.01 (**), and 0.001 level (***). Comparisons are between 1) male *T. scitulus* and male *T. polytmus*, 2) male and female *T. scitulus*, 3) female *T. scitulus* and female *T. polytmus*, and 4) male and female *T. polytmus*.

(A)

| | Degrees of Freedom | Sum Squares | Mean Squares | F | *p*-value | Significance |
|---|---|---|---|---|---|---|
| Group | 3 | 377.4 | 125.8 | 283.5 | 2.00E-16 | *** |
| Residuals | 149 | 66.1 | 0.44 | | | |

(B)

| Comparison | Difference | Lower | Upper | Adj. *p*-value | Significance |
|---|---|---|---|---|---|
| 1) Intraspecific *T. scitulus* | -3.190 | -3.683 | -2.698 | 0.000 | *** |
| 2) Intraspecific *T. polytmus* | -3.382 | -3.753 | -3.011 | 0.000 | *** |
| 3) Interspecific Male | -0.178 | -0.526 | 0.169 | 0.544 | NS |
| 4) Interspecific Female | 0.013 | -0.496 | 0.522 | 1.000 | NS |

Table S3.5. Per-locus gene diversities (Nei 1973) and allele numbers in six *Trochilus*-specific microsatellite markers for 171 individuals of *Trochilus polytmus*, *Trochilus scitulus*, and putative hybrids for 12 sites.

| | Site 1 | Site 2 | Site 3 | Site 4 | Site 5 | Site 6 | Site 7 | Site 8 | Site 9 | Site 10 | Site 11 | Site 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tro17 | 0.647/4 | 0.274/3 | 0.572/4 | 0.472/2 | 0.493/5 | 0.637/5 | 0.577/6 | 0.494/6 | 0.363/3 | 0.278/3 | 0.527/4 | 0.65/5 |
| Tro5 | 0.894/10 | 0.92/9 | 0.864/9 | 0.903/8 | 0.785/7 | 0.852/9 | 0.846/9 | 0.888/14 | 0.893/9 | 0.844/8 | 0.91/9 | 0.85/9 |
| Tro19 | 0.468/5 | 0.705/4 | 0.633/4 | 0/1 | 0.631/4 | 0.566/5 | 0.652/6 | 0.488/8 | 0.605/5 | 0.618/5 | 0.111/2 | 0.433/4 |
| Tro6 | 0.862/9 | 0.848/7 | 0.871/8 | 0.905/7 | 0.881/7 | 0.907/10 | 0.807/8 | 0.824/9 | 0.79/8 | 0.861/7 | 0.792/6 | 0.903/9 |
| Tro2 | 0.386/2 | 0.321/2 | 0.427/2 | 0.2/2 | 0.348/3 | 0.354/2 | 0.169/3 | 0.611/4 | 0.577/3 | 0.55/3 | 0.5/2 | 0.333/2 |
| Tro3 | 0.718/4 | 0.777/5 | 0.784/5 | 0.75/4 | 0.759/4 | 0.676/4 | 0.746/5 | 0.734/5 | 0.761/5 | 0.694/4 | 0.65/4 | 0.672/3 |

Table S3.6. AMOVA Table for six *Trochilus*-specific microsatellite markers. There is a small, yet significant, percentage of genetic variation among sites ($\Phi_{PT} = 0.051$, $P = 0.001$).

| Source | Degrees of Freedom | Sum Squares | Mean Squares | Estimated Variance | Percentage |
|---|---|---|---|---|---|
| Among Sites | 11 | 84.890 | 7.717 | 0.237 | 5% |
| Within Sites | 159 | 705.104 | 4.435 | 4.435 | 95% |
| Total | 170 | 789.994 | | 4.672 | 100% |

Table S3.7. Allele frequency differences in reference sites for 23 informative SNP loci (46 alleles) with DAPC allelic loadings above the threshold of 0.001. The absolute difference in allele frequencies between reference parental sites is reported.

| Allele | *Reference T. scitulus* | *Reference T. polytmus* | Abs. difference |
|---|---|---|---|
| LTP58155.002 | 0.1 | 0.7 | 0.6 |
| LTP58155.001 | 0.9 | 0.3 | 0.6 |
| LTP2447.001 | 0.4 | 0.9 | 0.5 |
| LTP2447.002 | 0.6 | 0.1 | 0.5 |
| LTP20160.002 | 0.3 | 0.8 | 0.5 |
| LTP20160.001 | 0.7 | 0.2 | 0.5 |
| LTP129531.002 | 0.1 | 0.6 | 0.5 |
| LTP129531.001 | 0.9 | 0.4 | 0.5 |
| LTP101621.002 | 0.2 | 0.7 | 0.5 |
| LTP101621.001 | 0.8 | 0.3 | 0.5 |
| LTP84654.002 | 0.4 | 0.8 | 0.4 |
| LTP84654.001 | 0.6 | 0.2 | 0.4 |
| LTP46068.002 | 0.0 | 0.4 | 0.4 |
| LTP46068.001 | 1.0 | 0.6 | 0.4 |
| LTP60803.001 | 0.6 | 0.3 | 0.4 |
| LTP60803.002 | 0.4 | 0.7 | 0.4 |
| LTP111732.001 | 0.2 | 0.5 | 0.3 |
| LTP111732.002 | 0.8 | 0.5 | 0.3 |
| LTP11923.002 | 0.5 | 0.8 | 0.3 |
| LTP11923.001 | 0.5 | 0.2 | 0.3 |
| LTP77030.002 | 0.3 | 0.1 | 0.3 |
| LTP77030.001 | 0.7 | 0.9 | 0.3 |
| LTP16537.002 | 0.8 | 0.5 | 0.3 |
| LTP16537.001 | 0.2 | 0.5 | 0.3 |
| LTP86688.002 | 0.6 | 0.9 | 0.3 |
| LTP98902.002 | 0.8 | 0.6 | 0.3 |
| LTP98902.001 | 0.2 | 0.4 | 0.3 |
| LTP86688.001 | 0.4 | 0.1 | 0.3 |
| LTP82072.001 | 0.8 | 0.6 | 0.2 |
| LTP82072.002 | 0.2 | 0.4 | 0.2 |
| LTP12516.002 | 0.4 | 0.3 | 0.2 |
| LTP12516.001 | 0.6 | 0.8 | 0.2 |
| LTP16188.002 | 0.4 | 0.6 | 0.2 |
| LTP16188.001 | 0.6 | 0.4 | 0.2 |
| LTP106396.002 | 0.5 | 0.3 | 0.2 |
| LTP23234.002 | 0.0 | 0.2 | 0.2 |
| LTP23234.001 | 1.0 | 0.8 | 0.2 |
| LTP106396.001 | 0.5 | 0.7 | 0.2 |

Table continued.

| Allele | *Reference T. scitulus* | *Reference T. polytmus* | Abs. difference |
|---|---|---|---|
| LTP86351.002 | 0.6 | 0.4 | 0.2 |
| LTP86351.001 | 0.4 | 0.6 | 0.2 |
| LTP101652.001 | 0.7 | 0.8 | 0.1 |
| LTP101652.002 | 0.3 | 0.2 | 0.1 |
| LTP97092.001 | 0.8 | 0.7 | 0.1 |
| LTP97092.002 | 0.2 | 0.3 | 0.1 |
| LTP41482.002 | 0.7 | 0.7 | 0.0 |
| LTP41482.001 | 0.3 | 0.3 | 0.0 |

**APPENDIX B.**

**SUPPLEMENTARY MATERIAL FOR CHAPTER 4**

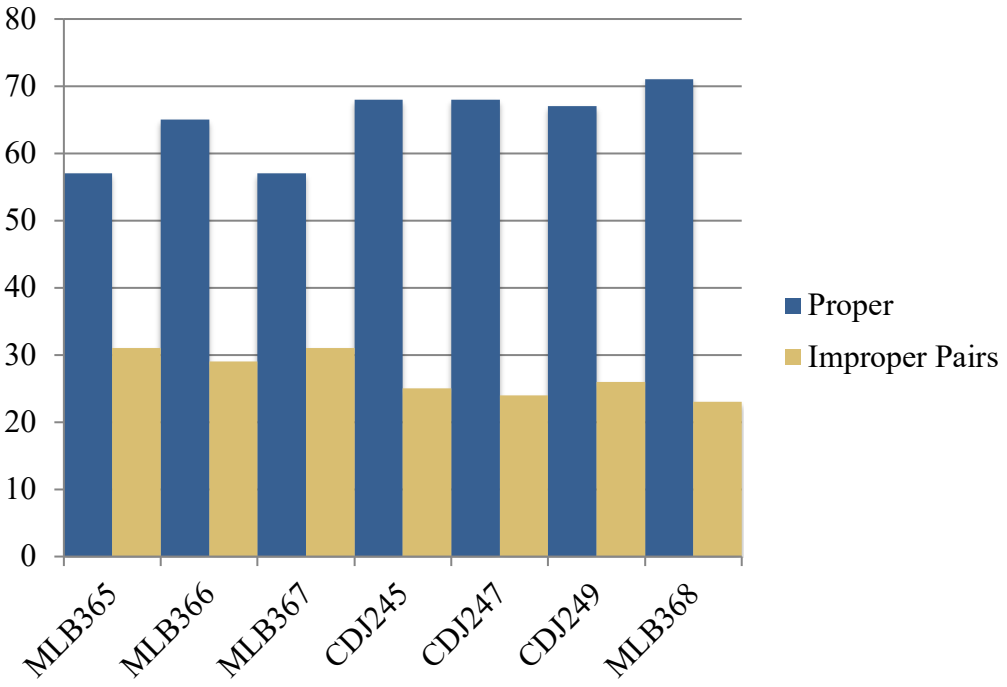



Fig. S4.1. The percentage of proper vs. improper pairs in seven libraries.

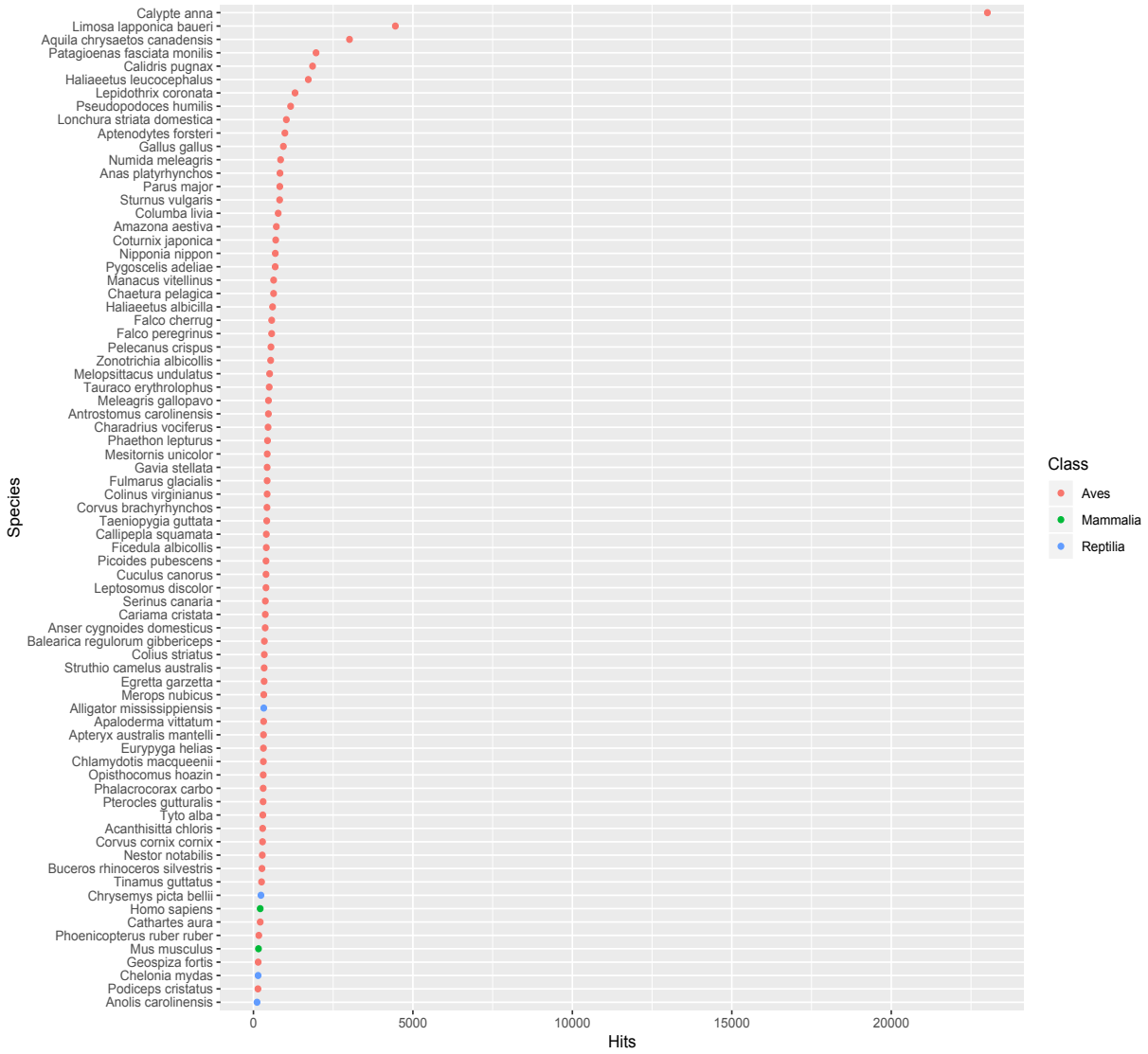



Fig. S4.2. The 75 species associated with the highest number of blast hits against the nr database using a minimum e-value of 1E-6. Only 6 of 75 are non-avian species.

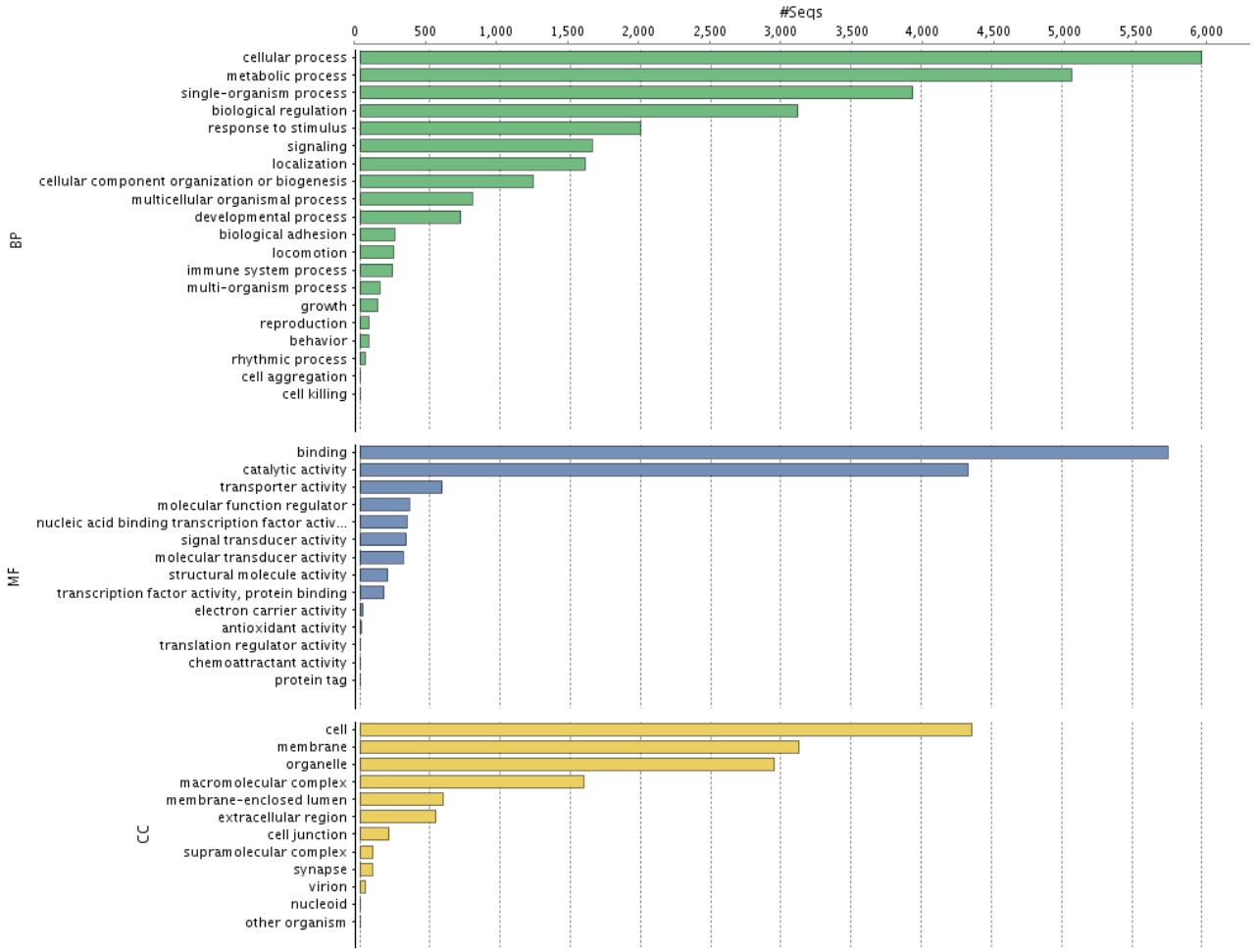



Fig. S4.3. The top 20 gene ontology (GO) terms mapped by by Blast2Go using blastx hits against NCBI's non-redundant protein database. GO terms are categorized by biological process (bp), molecular function (mf), and cellular component (cc).

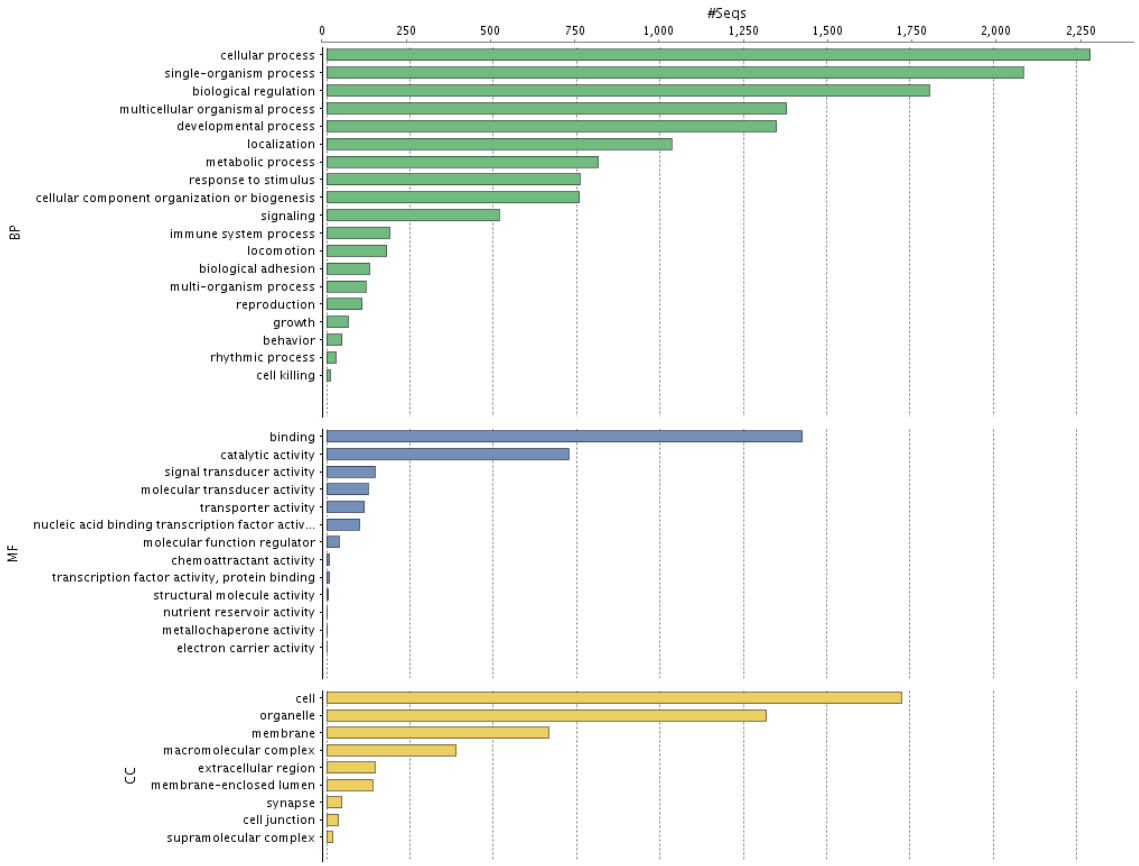



Fig. S4.4. The top 20 gene ontology (GO) terms mapped by Blast2Go using blastx hits against a custom-made candidate melanin gene database. GO terms are categorized by biological process (bp), molecular function (mf), and cellular component (cc).

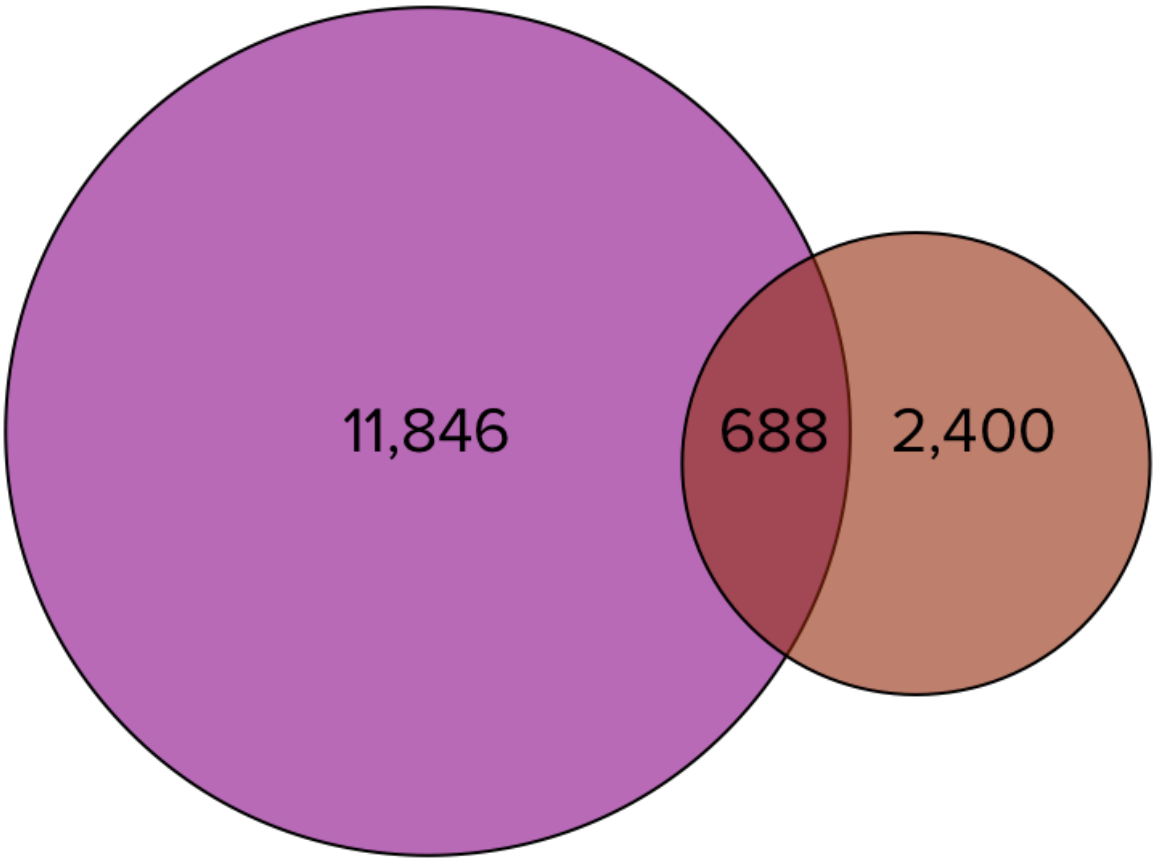



Fig. S4.5. Venn Diagram showing of annotated transcripts by Blast2Go using searches against the nr database (purple), and the pigmentation gene database (brown). Area of overlap represents transcripts recovered in both databases.

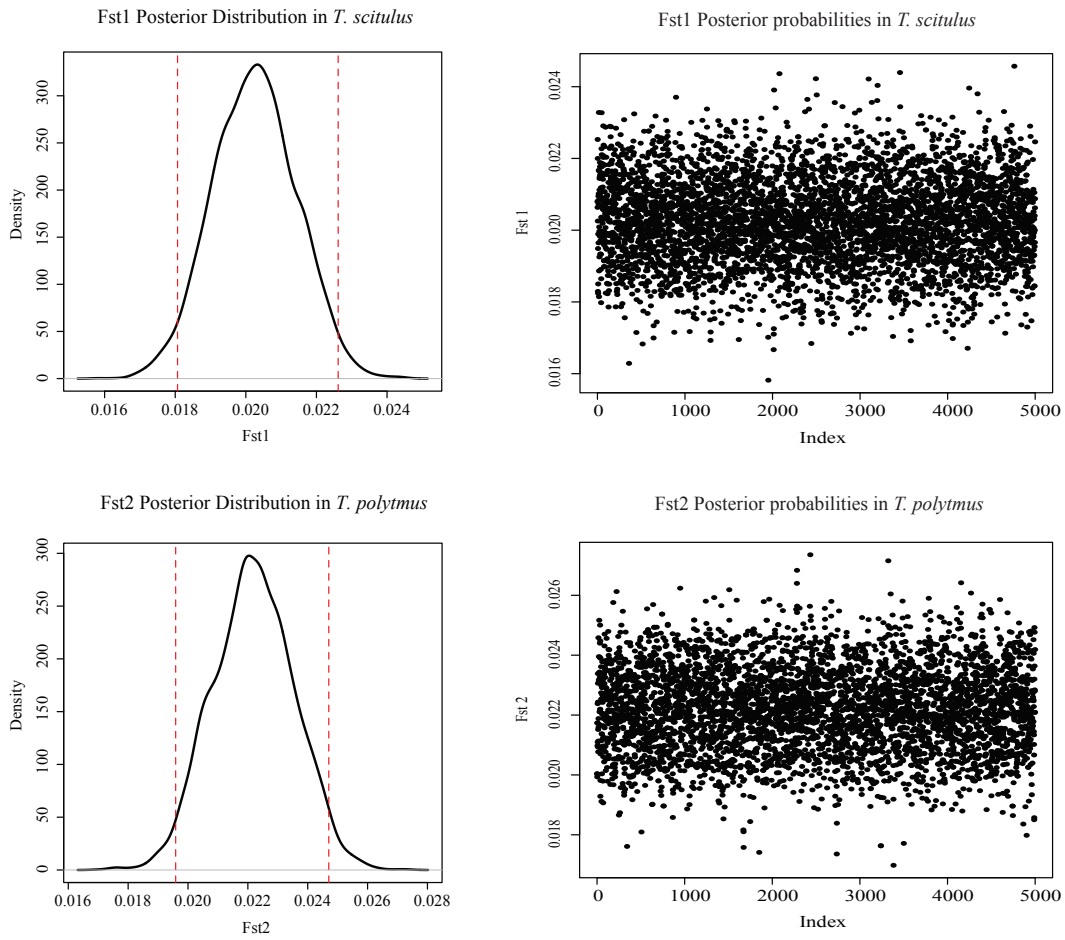



Fig. S4.6. Bayescan distribution of posterior values for $F_{ST}$ coefficients in *T. polytmus* and *T. scitulus*.

Table S4.1. Summary of trimming report from Trim_Galore! Library and field ID are presented along with the total number of reads processed, the percentage of reads in which adapters were detected, the number of reads written (passing filters), the number of base pairs that were processed, of those, the number that were trimmed off due to low quality, and the total number of base pairs written (passing filters). Finally, the number of pairs that were removed due to one or both reads being shorter than the 20 base pair threshold is reported.

| Library | Field ID | Total reads processed | Reads with adapters | Reads written | Total base pairs processed | Quality-trimmed | Total written (filtered) | Pairs removed |
|---|---|---|---|---|---|---|---|---|
| 2992 R1 | CDJ_238 | 41,646,381 | 32,365,602 (77.7%) | 41,646,381 (100.0%) | 6,288,603,531 | 1,165,590,158 (18.5%) | 2,653,737,344 (42.2%) | 24,014,755 (**57.66%**) |
| 2992 R2 | CDJ_238 | 41,646,381 | 32,120,352 (77.1%) | 41,646,381 (100.0%) | 6,288,603,531 | 1,657,451,456 (26.4%) | 2,536,157,267 (40.3%) | |
| 2993 R1 | CVD_245 | 31,192,389 | 11,244,559 (36.0%) | 31,192,389 (100.0%) | 4,710,050,739 | 101,131,257 (2.1%) | 4,437,245,601 (94.2%) | 1,502,540 (**4.82%**) |
| 2993 R2 | CVD_245 | 31,192,389 | 10,820,871 (34.7%) | 31,192,389 (100.0%) | 4,710,050,739 | 418,937,519 (8.9%) | 4,127,812,558 (87.6%) | |
| 2994 R1 | CVD_247 | 42,526,219 | 14,935,488 (35.1%) | 42,526,219 (100.0%) | 6,421,459,069 | 84,364,693 (1.3%) | 6,140,644,191 (95.6%) | 1,284,787 (**3.02%**) |
| 2994 R2 | CVD_247 | 42,526,219 | 14,195,284 (33.4%) | 42,526,219 (100.0%) | 6,421,459,069 | 510,926,313 (8.0%) | 5,724,904,881 (89.2%) | |
| 2995 R1 | CVD_249 | 31,218,859 | 11,775,802 (37.7%) | 31,218,859 (100.0%) | 4,714,047,709 | 158,301,579 (3.4%) | 4,292,435,053 (91.1%) | 2,579,943 (**8.26%**) |
| 2995 R2 | CVD_249 | 31,218,859 | 11,210,407 (35.9%) | 31,218,859 (100.0%) | 4,714,047,709 | 503,370,854 (10.7%) | 3,959,131,277 (84.0%) | |
| 2996 R1 | MLB_365 | 43,809,429 | 15,867,589 (36.2%) | 43,809,429 (100.0%) | 6,615,223,779 | 118,465,186 (1.8%) | 6,239,840,226 (94.3%) | 1,931,849 (**4.41%**) |

Table continued.

| Library | Field ID | Total reads processed | Reads with adapters | Reads written | Total base pairs processed | Quality-trimmed | Total written (filtered) | Pairs removed |
|---|---|---|---|---|---|---|---|---|
| 2996 R2 | MLB_365 | 43,809,429 | 15,135,014 (34.5%) | 43,809,429 (100.0%) | 6,615,223,779 | 556,690,084 (8.4%) | 5,815,595,407 (87.9%) | |
| 2997 R1 | MLB_366 | 27,943,516 | 10,459,187 (37.4%) | 27,943,516 (100.0%) | 4,219,470,916 | 76,287,305 (1.8%) | 3,966,355,407 (94.0%) | 1,246,020 (**4.46%**) |
| 2997 R2 | MLB_366 | 27,943,516 | 9,942,145 (35.6%) | 27,943,516 (100.0%) | 4,219,470,916 | 357,014,488 (8.5%) | 3,693,383,672 (87.5%) | |
| 2998 R1 | MLB_367 | 45,448,617 | 16,268,033 (35.8%) | 45,448,617 (100.0%) | 6,862,741,167 | 113,328,878 (1.7%) | 6,513,842,168 (94.9%) | 1,757,856 (**3.87%**) |
| 2998 R2 | MLB_367 | 45,448,617 | 15,750,481 (34.7%) | 45,448,617 (100.0%) | 6,862,741,167 | 540,593,197 (7.9%) | 6,098,495,144 (88.9%) | |
| 2999 R1 | MLB_368 | 41,743,217 | 13,647,831 (32.7%) | 41,743,217 (100.0%) | 6,303,225,767 | 505,031,548 (8.0%) | 5,623,207,733 (89.2%) | 1,488,198 (**3.57%**) |
| 2999 R2 | MLB_368 | 41,743,217 | 14,329,941 (34.3%) | 41,743,217 (100.0%) | 6,303,225,767 | 92,740,630 (1.5%) | 6,025,219,458 (95.6%) | |

Table S4.2. Assessment of assembly quality using contig N50 statistics. I report the contig N50 based on all transcripts, and the longest isoform per gene.

| Contig Stat | All Transcripts | Longest Isoform per Gene |
|---|---|---|
| Contig N10 | 4,768 | 3,699 |
| Contig N20 | 3,336 | 2,312 |
| Contig N30 | 2,480 | 1,550 |
| Contig N40 | 1,848 | 1,090 |
| Contig N50 | 1,360 | 814 |

Table S4.3. ExN50 Statistics. The minimum expression, E-N50, and number of transcripts are given for each E-N50 value.

| #E | Min. Exp. | E-N50 | Num. Transcripts |
|---|---|---|---|
| E24 | 355,540 | 1,587 | 1 |
| E42 | 270,213 | 1,587 | 2 |
| E52 | 191,234 | 1,196 | 3 |
| E61 | 187,119 | 1,587 | 4 |
| E64 | 40,823 | 1,531 | 5 |
| E67 | 32,340 | 1,531 | 6 |
| E69 | 29,519 | 1,587 | 7 |
| E70 | 17,144 | 1,587 | 8 |
| E71 | 12,418 | 1,587 | 9 |
| E72 | 10,911 | 1,644 | 10 |
| E73 | 7,244 | 1,786 | 12 |
| E74 | 4,923 | 1,786 | 14 |
| E75 | 3,629 | 1,644 | 18 |
| E76 | 2,544 | 1,637 | 23 |
| E77 | 2,043 | 1,531 | 28 |
| E78 | 1,654 | 1,437 | 35 |
| E79 | 1,596 | 1,531 | 42 |
| E80 | 1,255 | 1,532 | 51 |
| E81 | 976 | 1,403 | 61 |
| E82 | 788 | 1,396 | 76 |
| E83 | 582 | 1,299 | 94 |
| E84 | 465 | 1,196 | 118 |
| E85 | 351 | 1,161 | 150 |
| E86 | 273 | 1,273 | 189 |
| E87 | 202 | 1,322 | 242 |
| E88 | 141 | 1,287 | 313 |
| E89 | 102 | 1,396 | 413 |
| E90 | 65 | 1,448 | 562 |
| E91 | 40 | 1,669 | 807 |
| E92 | 23 | 1,757 | 1199 |
| E93 | 14 | 1,845 | 1859 |
| E94 | 8 | 1,961 | 2996 |
| E95 | 5 | 2,137 | 5054 |
| E96 | 2 | 2,263 | 8923 |
| E97 | 1 | 2,351 | 17191 |
| E98 | 0 | 2,206 | 36161 |
| E99 | 0 | 1,844 | 73255 |
| E100 | 0 | 1,455 | 162824 |

Table S4.4 Assessment of assembly quality using the number of full length coding sequences recovered by Blastx against the SWISSPROT database. The last column is the cumulative count of transcripts. For example, 7,426 transcripts contain greater than 80% of the protein sequence length.

| Percent Coverage Bin | Count in Bin | Cumulative Count |
|---|---|---|
| 100 | 5,233 | 5,233 |
| 90 | 1264 | 6,497 |
| 80 | 929 | 7,426 |
| 70 | 945 | 8,371 |
| 60 | 1117 | 9,488 |
| 50 | 1200 | 10,688 |
| 40 | 1342 | 12,030 |
| 30 | 1558 | 13,588 |
| 20 | 1580 | 15,168 |
| 10 | 716 | 15,884 |

Table S4.5. The GO terms discovered using the AMIGO2 browser used to search for relevant pigmentation genes.

| GO Term | Description |
|---|---|
| GO:0006583 | melanin biosynthetic process from tyrosine |
| GO:0006582 | melanin metabolic process |
| GO:0030354 | melanin-concentrating hormone activity |
| GO:0030273 | melanin-concentrating hormone receptor activity |
| GO:0042438 | melanin biosynthetic process |
| GO:0046150 | melanin catabolic process |
| GO:0031778 | type 2 melanin-concentrating hormone receptor binding |
| GO:0031776 | melanin-concentrating hormone receptor binding |
| GO:0031777 | type 1 melanin-concentrating hormone receptor binding |
| GO:0048023 | positive regulation of melanin biosynthetic process |
| GO:0048022 | negative regulation of melanin biosynthetic process |
| GO:0048021 | regulation of melanin biosynthetic process |
| GO:0031409 | pigment binding |
| GO:0043324 | pigment metabolic process involved in developmental pigmentation |
| GO:0097324 | melanocyte migration |
| GO:0097325 | melanocyte proliferation |
| GO:0042470 | melanosome |
| GO:0003406 | retinal pigment epithelium development |
| GO:0042440 | pigment metabolic process |
| GO:0032438 | melanosome organization |
| GO:0046149 | pigment catabolic process |
| GO:0046148 | pigment biosynthetic process |
| GO:1903232 | melanosome assembly |
| GO:0036485 | dorsolateral trunk neural crest cell migration |
| GO:0035006 | melanization defense response |
| GO:0036160 | melanocyte-stimulating hormone secretion |
| GO:0075043 | maintenance of turgor in appressorium by melanization |
| GO:0042470 | melanosome |
| GO:0035646 | endosome to melanosome transport |
| GO:1902908 | regulation of melanosome transport |
| GO:1902909 | negative regulation of melanosome transport |
| GO:1902910 | positive regulation of melanosome transport |
| GO:0034493 | melanosome lumen |
| GO:0032401 | establishment of melanosome localization |
| GO:0032402 | melanosome transport |
| GO:0032400 | melanosome localization |
| GO:0032438 | melanosome organization |
| GO:1903056 | regulation of melanosome organization |
| GO:1903057 | negative regulation of melanosome organization |

Table continued.

| GO Term | Description |
|---|---|
| GO:1903058 | positive regulation of melanosome organization |
| GO:0033162 | melanosome membrane |
| GO:1903232 | melanosome assembly |
| GO:0017044 | melanocyte-stimulating hormone activity |
| GO:1990680 | response to melanocyte-stimulating hormone |
| GO:0035656 | kinesin-associated melanosomal adaptor activity |
| GO:0031084 | *BLOC-2* complex |
| GO:0031085 | *BLOC-3* complex |
| GO:0031082 | *BLOC* complex |
| GO:0031083 | *BLOC-1* complex |
| GO:0008502 | melatonin receptor activity |
| GO:0036461 | *BLOC-2* complex binding |

Table S4.6. List of 486 accessions of melanin-based pigmentation genes in the custom database.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| ANKRD27 | H0ZFM1 | 100223380 | *Taeniopygia guttata* | Ankyrin repeat domain 27 |
| ANKRD27 | R4GL41 | 415766 | *Gallus gallus* | Ankyrin repeat domain 27 |
| ANKRD27 | U3IWN8 | 101790732 | *Anas platyrhynchos* | Ankyrin repeat domain 27 |
| ANKRD27 | U3JBV6 | 101806832 | *Ficedula albicollis* | Ankyrin repeat domain 27 |
| ANXA6 | P51901 | 395481 | *Gallus gallus* | Annexin A6 |
| AP1G1 | H0ZDB3 | 100217944 | *Taeniopygia guttata* | Adaptor related protein complex 1 gamma 1 subunit |
| AP1G1 | U3J3Q9 | 101790053 | *Anas platyrhynchos* | Adaptor related protein complex 1 gamma 1 subunit |
| AP1G1 | U3JIW1 | 101817419 | *Ficedula albicollis* | Adaptor related protein complex 1 gamma 1 subunit |
| AP1M1 | H0YR12 | 100227179 | *Taeniopygia guttata* | Adaptor related protein complex 1 mu 1 subunit |
| AP1M1 | Q5ZMG7 | 420149 | *Gallus gallus* | Adaptor related protein complex 1 mu 1 subunit |
| AP1M1 | U3INH6 | 101797387 | *Anas platyrhynchos* | Adaptor related protein complex 1 mu 1 subunit |
| AP1M1 | U3KG68 | 101808882 | *Ficedula albicollis* | Adaptor related protein complex 1 mu 1 subunit |
| AP3B1 | E1BW97 | 427646 | *Gallus gallus* | AP-3 complex subunit beta |
| AP3B1 | H0Z0I6 | 100226392 | *Taeniopygia guttata* | AP-3 complex subunit beta |
| AP3D1 | G1MTH1 | 100546834 | *Meleagris gallopavo* | Adaptor related protein complex 3 delta 1 subunit |
| AP3D1 | H0YPU7 | 100228101 | *Taeniopygia guttata* | Adaptor related protein complex 3 delta 1 subunit |
| AP3D1 | U3IQE6 | 101790870 | *Anas platyrhynchos* | AP-3 complex subunit delta |
| ASIP | G1MTH0 | 100542191 | *Meleagris gallopavo* | Agouti signaling protein |
| ASIP | U3K7P9 | 101806912 | *Ficedula albicollis* | Agouti signaling protein |
| ATP6V0A1 | Q9I8D0 | 395474 | *Gallus gallus* | V-type proton ATPase 116 kDa subunit a isoform 1 |
| BLOC1S2 | G1N6K3 | 100545527 | *Meleagris gallopavo* | Biogenesis of lysosomal organelles complex 1S2 |
| BLOC1S4 | F1NWV3 | 421000 | *Gallus gallus* | Biogenesis of lysosomal organelles complex 1S4 (cpo) |
| BLOC1S4 | H0ZEZ2 | 100229155 | *Taeniopygia guttata* | Biogenesis of lysosomal organelles complex 1S4 |
| BLOC1S5 | A0A091RU56 | 104537634 | *Mesitornis unicolor* | Biogenesis of lysosomal organelles complex 1S5 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| BLOC1S5 | A0A093G7D0 | 104308915 | *Picoides pubescens* | Biogenesis of lysosomal organelles complex 1S5 |
| BLOC1S5 | Q5ZK77 | 420866 | *Gallus gallus* | Biogenesis of lysosomal organelles complex 1S5 |
| BLOC1S5 | H0YVW5 | 100227726 | *Taeniopygia guttata* | Biogenesis of lysosomal organelles complex 1S5 |
| BLOC1S6 | Q5ZIM2 | 415447 | *Gallus gallus* | Biogenesis of lysosomal organelles complex 1S6 (Pldn) |
| BMP4 | A5HMF8 | 101803077 | *Anas platyrhynchos* | Bone morphogenetic protein 4 |
| BSG | P17790 | 770363 | *Gallus gallus* | Basigin |
| CDH3 | U3JK13 | 101805798 | *Ficedula albicollis* | Cadherin 3 |
| CNP | H0YW77 | 100223675 | *Taeniopygia guttata* | 2,3-cyclic-nucleotide 3-phosphodiesterase |
| CNP | O57389 | 395921 | *Gallus gallus* | 2,3-cyclic-nucleotide 3-phosphodiesterase |
| DCT | O93505 | 395775 | *Gallus gallus* | L-dopachrome tautomerase |
| DCT | H0ZLV3 | 100230498 | *Taeniopygia guttata* | Dopachrome tautomerase |
| DCT | U3J2E9 | 101792699 | *Anas platyrhynchos* | Dopachrome tautomerase |
| DCT | U3JPL0 | 101812075 | *Ficedula albicollis* | Dopachrome tautomerase |
| DTNBP1 | G1MVC8 | 100549563 | *Meleagris gallopavo* | Dystrobrevin binding protein 1 |
| DTNBP1 | G3UV03 | 100549563 | *Meleagris gallopavo* | Dystrobrevin binding protein 1 |
| DTNBP1 | U3KG51 | 101819807 | *Ficedula albicollis* | Dystrobrevin binding protein 1 |
| DTNBP1 | Q5ZKM0 | 420840 | *Gallus gallus* | Dysbindin |
| FAP | F1NDP3 | 424186 | *Gallus gallus* | Uncharacterized protein |
| FAP | H0Z954 | 100228501 | *Taeniopygia guttata* | Fibroblast activation protein alpha |
| FAP | U3JD75 | 101808860 | *Ficedula albicollis* | Fibroblast activation protein alpha |
| GCHFR | G1MVC2 | 100540224 | *Meleagris gallopavo* | GTP cyclohydrolase I feedback regulator |
| GCHFR | H0Z8D6 | 100220034 | *Taeniopygia guttata* | GTP cyclohydrolase I feedback regulator |
| GCHFR | U3KB51 | 101809670 | *Ficedula albicollis* | GTP cyclohydrolase I feedback regulator |
| GPR143 | E1BRL2 | 418652 | *Gallus gallus* | G protein-coupled receptor 143 |
| GPR143 | G1NPE8 | 100545250 | *Meleagris gallopavo* | G protein-coupled receptor 143 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| GPR143 | H0ZDT4 | 100224689 | *Taeniopygia guttata* | G protein-coupled receptor 143 |
| GPR143 | U3IDB0 | 101795267 | *Anas platyrhynchos* | G protein-coupled receptor 143 |
| GPR143 | U3JPJ9 | 101806142 | *Ficedula albicollis* | G protein-coupled receptor 143 |
| GPR50 | F1NEF4 | 396318 | *Gallus gallus* | Uncharacterized protein |
| GPR50 | P49288 | 396318 | *Gallus gallus* | Melatonin receptor type 1C |
| HPS1 | H0ZGW7 | 100222793 | *Taeniopygia guttata* | HPS1, biogenesis of lysosomal organelles complex 3S1 |
| HPS1 | Q5ZIL0 | 429879 | *Gallus gallus* | HPS1, biogenesis of lysosomal organelles complex 3S1 |
| HPS1 | U3IX30 | 101793973 | *Anas platyrhynchos* | HPS1, biogenesis of lysosomal organelles complex 3S1 |
| HPS3 | G1NEX7 | 100549787 | *Meleagris gallopavo* | HPS3, biogenesis of lysosomal organelles complex 2S1 |
| HPS3 | H0ZMI5 | 100232307 | *Taeniopygia guttata* | HPS3, biogenesis of lysosomal organelles complex 2S1 |
| HPS3 | U3IHW9 | 101789779 | *Anas platyrhynchos* | HPS3, biogenesis of lysosomal organelles complex 2S1 |
| HPS4 | F1NQT3 | 416907 | *Gallus gallus* | HPS4, biogenesis of lysosomal organelles complex 3S2 |
| HPS4 | G1N533 | 100542291 | *Meleagris gallopavo* | HPS4, biogenesis of lysosomal organelles complex 3S2 |
| HPS5 | G1N610 | 100549976 | *Meleagris gallopavo* | HPS5, biogenesis of lysosomal organelles complex 2S2 |
| HPS5 | H0ZEZ8 | 100222927 | *Taeniopygia guttata* | HPS5, biogenesis of lysosomal organelles complex 2S2 |
| HPS5 | U3ICC4 | 101801874 | *Anas platyrhynchos* | HPS5, biogenesis of lysosomal organelles complex 2S2 |
| HPS5 | U3K9B1 | 101821001 | *Ficedula albicollis* | HPS5, biogenesis of lysosomal organelles complex 2S2 |
| HPS6 | E1C2Y8 | 770328 | *Gallus gallus* | HPS6, biogenesis of lysosomal organelles complex 2S3 |
| HSP90AA1 | P11501 | 423463 | *Gallus gallus* | Heat shock protein HSP 90-alpha |
| HSP90AB1 | Q04619 | 396188 | *Gallus gallus* | Heat shock cognate protein HSP 90-beta |
| IHH | H9KZS8 | 395801 | *Gallus gallus* | Hedgehog protein |
| IHH | Q98938 | 395801 | *Gallus gallus* | Indian hedgehog protein C-product |
| KIF13A | F1P140 | 420832 | *Gallus gallus* | Kinesin family member 13A |
| KIF13A | G1MUW7 | 100547928 | *Meleagris gallopavo* | Kinesin family member 13A |
| KIF13A | U3ICL4 | 101794022 | *Anas platyrhynchos* | Kinesin family member 13A |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| KIF13A | U3KGB9 | 101820984 | *Ficedula albicollis* | Kinesin family member 13A |
| KIT | G1NH11 | 100542786 | *Meleagris gallopavo* | KIT proto-oncogene receptor tyrosine kinase |
| KIT | H0ZBL7 | 100221238 | *Taeniopygia guttata* | KIT proto-oncogene receptor tyrosine kinase |
| KIT | Q08156 | 378783 | *Gallus gallus* | Mast/stem cell growth factor receptor Kit |
| KIT | U3I8W5 | 101796083 | *Anas platyrhynchos* | KIT proto-oncogene receptor tyrosine kinase |
| KXD1 | H0YS17 | 100229067 | *Taeniopygia guttata* | Uncharacterized protein |
| LAMP1 | G3UUP0 | 100548531 | *Meleagris gallopavo* | Lysosomal associated membrane protein 1 |
| LAMP1 | U3J7H8 | 101795078 | *Anas platyrhynchos* | Lysosomal associated membrane protein 1 |
| LOC100190527 | B5G4H7 | 100190527 | *Taeniopygia guttata* | Putative snapin |
| LOC100218322 | H0Z6L6 | 100218322 | *Taeniopygia guttata* | Dystrobrevin binding protein 1 |
| LOC100219393 | H0ZBN4 | 100219393 | *Taeniopygia guttata* | Biogenesis of lysosomal organelles complex 1 subunit 2 |
| LOC100539772 | G1NE37 | 100539772 | *Meleagris gallopavo* | Uncharacterized protein |
| LOC100547273 | G1NEL7 | 100547273 | *Meleagris gallopavo* | Fibroblast activation protein alpha |
| LOC100549505 | G1NRE4 | 100549505 | *Meleagris gallopavo* | Uncharacterized protein |
| LOC100550272 | G1NPI8 | 100550272 | *Meleagris gallopavo* | OCA2 melanosomal transmembrane protein |
| LOC101799623 | U3ISJ2 | 101799623 | *Anas platyrhynchos* | Fibroblast activation protein alpha |
| MED1 | E1BUI5 | 420004 | *Gallus gallus* | Mediator of RNA polymerase II transcription subunit 1 |
| MED1 | U3J0Z2 | 101795501 | *Anas platyrhynchos* | Mediator of RNA polymerase II transcription subunit 1 |
| MLANA | E1C403 | 769648 | *Gallus gallus* | Melan-A |
| MLANA | U3I7Z8 | 101797582 | *Anas platyrhynchos* | Melan-A |
| MLANA | U3K7R0 | 101810085 | *Ficedula albicollis* | Melan-A |
| MREG | E1BQS6 | 424013 | *Gallus gallus* | Melanoregulin |
| MREG | H0Z094 | 100231672 | *Taeniopygia guttata* | Melanoregulin |
| MTNR1A | A0A093GEJ9 | 104301790 | *Picoides pubescens* | Melatonin receptor type 1A |
| MTNR1A | Q0ZAL0 | 751586 | *Taeniopygia guttata* | Mel-1a melatonin receptor |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| MTNR1A | U3K9C9 | 101818649 | *Ficedula albicollis* | Melatonin receptor 1A |
| MTNR1B | Q3LTL8 | 751588 | *Taeniopygia guttata* | Mel-1b melatonin receptor |
| MTNR1B | U3JBY1 | 101813273 | *Ficedula albicollis* | Uncharacterized protein |
| MYO5A | H0Z8L0 | 100225447 | *Taeniopygia guttata* | Myosin VA |
| MYO5A | U3I653 | 101798412 | *Anas platyrhynchos* | Myosin VA |
| MYO7A | G1NQT3 | 100543555 | *Meleagris gallopavo* | Myosin VIIA |
| MYO7A | U3IJE5 | 101796115 | *Anas platyrhynchos* | Myosin VIIA |
| MYRIP | G1NGS3 | 100543829 | *Meleagris gallopavo* | Myosin VIIA and Rab interacting protein |
| MYRIP | H0Z2L2 | 100221630 | *Taeniopygia guttata* | Myosin VIIA and Rab interacting protein |
| OCA2 | F1NES9 | 428009 | *Gallus gallus* | P protein |
| OCA2 | H0ZJL2 | 100228502 | *Taeniopygia guttata* | OCA2 melanosomal transmembrane protein |
| OCA2 | R0LB71 | 101795922 | *Anas platyrhynchos* | P protein |
| PDIA3 | Q8JG64 | 373899 | *Gallus gallus* | Protein disulfide-isomerase A3 |
| PMEL | Q98917 | 396007 | *Gallus gallus* | Melanocyte protein PMEL |
| PPIB | P24367 | 396447 | *Gallus gallus* | Peptidyl-prolyl cis-trans isomerase B |
| RAB11A | G1NBU1 | 100542957 | *Meleagris gallopavo* | RAB11A, member RAS oncogene family |
| RAB11A | Q5ZJN2 | 415544 | *Gallus gallus* | Ras-related protein Rab-11A |
| RAB11B | H0YP88 | 100220367 | *Taeniopygia guttata* | RAB11B, member RAS oncogene family |
| RAB11B | U3IU57 | 101796061 | *Anas platyrhynchos* | RAB11B, member RAS oncogene family |
| RAB11B | U3KBE4 | 101812004 | *Ficedula albicollis* | RAB11B, member RAS oncogene family |
| RAB17 | E1C6J7 | 424017 | *Gallus gallus* | RAB17, member RAS oncogene family |
| RAB17 | H0Z014 | 100218317 | *Taeniopygia guttata* | RAB17, member RAS oncogene family |
| RAB17 | U3KE91 | 101810876 | *Ficedula albicollis* | RAB17, member RAS oncogene family |
| RAB1A | G1MZK2 | 100538875 | *Meleagris gallopavo* | RAB1A, member RAS oncogene family |
| RAB1A | U3IPY5 | 101790987 | *Anas platyrhynchos* | RAB1A, member RAS oncogene family |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| RAB1A | U3KGR4 | 101807437 | *Ficedula albicollis* | RAB1A, member RAS oncogene family |
| RAB27A | D2D3P4 | 415410 | *Gallus gallus* | Rab27a |
| RAB27A | H0Z802 | 100230193 | *Taeniopygia guttata* | RAB27A, member RAS oncogene family |
| RAB27A | R0M2P9 | 101797215 | *Anas platyrhynchos* | Ras-related protein Rab-27A |
| RAB27A | U3KBY1 | 101810061 | *Ficedula albicollis* | RAB27A, member RAS oncogene family |
| RAB32 | F1NBL3 | 421616 | *Gallus gallus* | RAB32, member RAS oncogene family |
| RAB32 | H0ZKG3 | 100224194 | *Taeniopygia guttata* | RAB32, member RAS oncogene family |
| RAB32 | U3KF61 | 101807563 | *Ficedula albicollis* | RAB32, member RAS oncogene family |
| RAB38 | H0ZRQ0 | 100227935 | *Taeniopygia guttata* | RAB38, member RAS oncogene family |
| RAB38 | R0L987 | 101804947 | *Anas platyrhynchos* | Ras-related protein Rab-38 |
| RAB38 | R4GLM3 | 428095 | *Gallus gallus* | RAB38, member RAS oncogene family |
| RAB9A | H0ZD10 | 100225146 | *Taeniopygia guttata* | RAB9A, member RAS oncogene family |
| RAB9A | Q5ZMI5 | 418635 | *Gallus gallus* | RAB9A, member RAS oncogene family |
| RAB9A | R0LG13 | 101804925 | *Anas platyrhynchos* | Ras-related protein Rab-9A |
| RAB9A | U3KL70 | 101818022 | *Ficedula albicollis* | RAB9A, member RAS oncogene family |
| RARB | U3IFR1 | 101793182 | *Anas platyrhynchos* | Retinoic acid receptor beta |
| RARB | U3K3J3 | 101815644 | *Ficedula albicollis* | Retinoic acid receptor beta |
| RARB | P22448 | 396266 | *Gallus gallus* | Retinoic acid receptor beta |
| SEC22B | Q5ZJW4 | 424377 | *Gallus gallus* | Vesicle-trafficking protein SEC22b |
| SGSM2 | F1N8F2 | 417673 | *Gallus gallus* | Small G protein signaling modulator 2 |
| SGSM2 | U3I723 | 101805400 | *Anas platyrhynchos* | Small G protein signaling modulator 2 |
| SGSM2 | U3JQ99 | 101816085 | *Ficedula albicollis* | Small G protein signaling modulator 2 |
| SHH | A0A091GJG3 | 104057035 | *Cuculus canorus* | Hedgehog protein |
| SHH | H0YR00 | 100222391 | *Taeniopygia guttata* | Hedgehog protein |
| SHH | Q91035 | 395615 | *Gallus gallus* | Sonic hedgehog protein N-product |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| SHROOM2 | A0A1D5P807 | 418651 | *Gallus gallus* | Shroom family member 2 |
| SLC24A5 | S4SXD8 | 101798864 | *Anas platyrhynchos* | Solute carrier family 24 member 5 |
| SLC24A5 | G1N503 | 100550043 | *Meleagris gallopavo* | Solute carrier family 24 member 5 |
| SLC24A5 | H0Z9S7 | 100229271 | *Taeniopygia guttata* | Solute carrier family 24 member 5 |
| SNAP25 | G1N3J4 | 100545824 | *Meleagris gallopavo* | Synaptosomal-associated protein |
| SNAP25 | H0Z6W7 | 751982 | *Taeniopygia guttata* | Synaptosomal-associated protein |
| SNAP25 | U3IMQ0 | 101800247 | *Anas platyrhynchos* | Synaptosomal-associated protein |
| SNAP25 | U3K932 | 101810042 | *Ficedula albicollis* | Synaptosomal-associated protein |
| SNAP47 | F1NX76 | 420399 | *Gallus gallus* | Synaptosome associated protein 47 |
| SNAP47 | U3IBF6 | 101801275 | *Anas platyrhynchos* | Synaptosome associated protein 47 |
| SNAP47 | U3JEE0 | 101812229 | *Ficedula albicollis* | Synaptosome associated protein 47 |
| SNAPIN | R4GG77 | 100858163 | *Gallus gallus* | SNARE-associated protein Snapin |
| STX12 | E1C319 | 769300 | *Gallus gallus* | Syntaxin 12 |
| STX12 | G1MQW4 | 100543272 | *Meleagris gallopavo* | Syntaxin 12 |
| STX12 | H0YRF7 | 100223777 | *Taeniopygia guttata* | Syntaxin 12 |
| STX3 | G1MWG3 | 100545263 | *Meleagris gallopavo* | Syntaxin 3 |
| SYTL2 | U3IW92 | 101804717 | *Anas platyrhynchos* | Synaptotagmin like 2 |
| TBC1D32 | F1N9K0 | 421725 | *Gallus gallus* | TBC1 domain family member 32 |
| TBC1D32 | U3J594 | 101794603 | *Anas platyrhynchos* | TBC1 domain family member 32 |
| TBC1D32 | U3K9Q3 | 101814530 | *Ficedula albicollis* | TBC1 domain family member 32 |
| TFRC | Q90997 | 396191 | *Gallus gallus* | Transferrin receptor protein 1 |
| TH | H0ZG94 | 100219119 | *Taeniopygia guttata* | Tyrosine hydroxylase |
| TH | Q9PU40 | 395592 | *Gallus gallus* | Tyrosine hydroxylase |
| TH | U3KAM3 | 101811971 | *Ficedula albicollis* | Tyrosine hydroxylase |
| TMEM33 | G1NHW1 | 100547723 | *Meleagris gallopavo* | Transmembrane protein 33 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| TMEM33 | U3IJS6 | 101801839 | *Anas platyrhynchos* | Transmembrane protein 33 |
| TMEM33 | U3JJR9 | 101811643 | *Ficedula albicollis* | Transmembrane protein 33 |
| TYR | G1NQP6 | 100538462 | *Meleagris gallopavo* | Tyrosinase |
| TYR | H0ZRN9 | 100225893 | *Taeniopygia guttata* | Tyrosinase |
| TYRP1 | H0Z2R8 | 100218715 | *Taeniopygia guttata* | Tyrosinase related protein 1 |
| VPS33B | Q5ZL71 | 425858 | *Gallus gallus* | VPS33B, late endosome and lysosome associated |
| WASHC1 | F1NMK3 | 418145 | *Gallus gallus* | WASH complex subunit 1 |
| WNT5A | H0Z9G4 | 100219241 | *Taeniopygia guttata* | Protein Wnt |
| WNT5A | Q9YGX6 | 395703 | *Gallus gallus* | Protein Wnt |
| WNT5A | R0L7T2 | 101802371 | *Anas platyrhynchos* | Protein Wnt |
| ZEB2 | A0A0U4VU53 | 424306 | *Gallus gallus* | Zinc finger E-box binding homeobox 2 |
| ZEB2 | H0ZNU3 | 100220433 | *Taeniopygia guttata* | Zinc finger E-box binding homeobox 2 |
| ZEB2 | U3IJB5 | 101792999 | *Anas platyrhynchos* | Zinc finger E-box binding homeobox 2 |
| ABRAXAS1 | G1N7W1 | 100545211 | *Meleagris gallopavo* | Abraxas 1, BRCA1 A complex subunit |
| ABRAXAS1 | G3UQ58 | 100545211 | *Meleagris gallopavo* | Abraxas 1, BRCA1 A complex subunit |
| ABRAXAS1 | U3IYM0 | 101798365 | *Anas platyrhynchos* | Abraxas 1, BRCA1 A complex subunit |
| ACD | A0A091CMC3 | 104854001 | *Fukomys damarensis* | Adrenocortical dysplasia protein like protein |
| ACD | A0A096NR79 | 101019163 | *Papio anubis* | ACD, shelterin complex subunit and telomerase recruitment factor |
| ACD | A0A0C4DGT6 | | *Homo sapiens* | Adrenocortical dysplasia protein homolog |
| ADAM17 | H0ZS26 | 100225922 | *Taeniopygia guttata* | ADAM metallopeptidase domain 17 |
| ADAM17 | Q5ZL93 | 421931 | *Gallus gallus* | Uncharacterized protein |
| ADAM17 | U3IJI0 | 101794295 | *Anas platyrhynchos* | ADAM metallopeptidase domain 17 |
| ADAM17 | U3JVW2 | 101814343 | *Ficedula albicollis* | ADAM metallopeptidase domain 17 |
| ADAMTS20 | A0A1D5P8Z5 | 417794 | *Gallus gallus* | ADAM metallopeptidase with thrombospondin 1M20 |
| ADAMTS20 | H0Z683 | 100221192 | *Taeniopygia guttata* | ADAM metallopeptidase with thrombospondin 1M20 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| ANK1 | F1NJR5 | 396311 | *Gallus gallus* | Ankyrin 1 |
| ANK1 | H0Z4N4 | 100221183 | *Taeniopygia guttata* | Ankyrin 1 |
| ANK1 | U3IF32 | 101792152 | *Anas platyrhynchos* | Ankyrin 1 |
| ARL6 | A0A091EK49 | 103613688 | *Corvus brachyrhynchos* | ADP-ribosylation factor-like 6 |
| ARL6 | A0A091JCD1 | 104127768 | *Egretta garzetta* | ADP-ribosylation factor-like 6 |
| ARL6 | A0A1L1RMR6 | 418433 | *Gallus gallus* | ADP ribosylation factor like GTPase 6 |
| ARL6 | H0ZTF1 | 100228152 | *Taeniopygia guttata* | ADP ribosylation factor like GTPase 6 |
| ASCL1 | Q90575 | 386573 | *Gallus gallus* | Achaete-scute homologue |
| ATOX1 | F1P2A1 | 770231 | *Gallus gallus* | Antioxidant 1 copper chaperone |
| ATP7A | G1N398 | 100550653 | *Meleagris gallopavo* | ATPase copper transporting alpha |
| ATP7A | H0Z6L2 | 100217932 | *Taeniopygia guttata* | Uncharacterized protein |
| ATP7A | U3IIB7 | 101794668 | *Anas platyrhynchos* | ATPase copper transporting alpha |
| ATP7A | U3K8S3 | 101811573 | *Ficedula albicollis* | ATPase copper transporting alpha |
| ATP7B | F1P5C8 | 418879 | *Gallus gallus* | ATPase copper transporting beta |
| ATP7B | G1NQ71 | 100541545 | *Meleagris gallopavo* | ATPase copper transporting beta |
| ATP7B | H0ZPA1 | 100219575 | *Taeniopygia guttata* | Uncharacterized protein |
| ATP7B | U3IFE2 | 101795264 | *Anas platyrhynchos* | ATPase copper transporting beta |
| ATRN | F1P2W2 | 422946 | *Gallus gallus* | Attractin |
| ATRN | G1NKN8 | 100540379 | *Meleagris gallopavo* | Attractin |
| ATRN | U3IBE8 | 101795507 | *Anas platyrhynchos* | Attractin |
| BABAM1 | A0A1D5PX35 | 776677 | *Gallus gallus* | Uncharacterized protein |
| BABAM1 | H9H0P7 | 100549297 | *Meleagris gallopavo* | Uncharacterized protein |
| BABAM2 | A0A091NWM5 | 103805648 | *Acanthisitta chloris* | BRCA1-A complex subunit BRE |
| BABAM2 | A0A099Z0R9 | 104565084 | *Tinamus guttatus* | BRCA1-A complex subunit BRE |
| BABAM2 | H0YWE9 | 100231620 | *Taeniopygia guttata* | BRISC and BRCA1 A complex member 2 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| BAP1 | A0A091ENQ7 | 103614276 | *Corvus brachyrhynchos* | Ubiquitin carboxyl-terminal hydrolase |
| BAP1 | Q5F3N6 | 415944 | *Gallus gallus* | Ubiquitin carboxyl-terminal hydrolase BAP1 |
| BAP1 | H0Z7A1 | 100224993 | *Taeniopygia guttata* | BRCA1 associated protein 1 |
| BARD1 | A0A091QI14 | 103769326 | *Merops nubicus* | BRCA1-associated RING domain protein 1 |
| BARD1 | H0YXG0 | 100229796 | *Taeniopygia guttata* | BRCA1 associated RING domain 1 |
| BARD1 | Q5ZKJ1 | 424010 | *Gallus gallus* | Uncharacterized protein |
| BARD1 | U3J7H0 | 101800768 | *Anas platyrhynchos* | BRCA1 associated RING domain 1 |
| BAX | A0A218U8I9 | | *Lonchura striata* | Apoptosis regulator BAX |
| BAX | H0ZZC9 | | *Taeniopygia guttata* | Uncharacterized protein |
| BCL2 | A0A091W0Q8 | 104021857 | *Nipponia nippon* | Apoptosis regulator Bcl-2 |
| BCL2 | A0A099Z058 | 104580966 | *Tinamus guttatus* | Apoptosis regulator Bcl-2 |
| BCL2 | H0YUX3 | 100224888 | *Taeniopygia guttata* | BCL2, apoptosis regulator |
| BLOC1S1 | D3YVM8 | | *Mus musculus* | Biogenesis of lysosomal organelles complex 1 subunit 1 |
| BLOC1S1 | F8VP73 | | *Homo sapiens* | Biogenesis of lysosomal organelles complex 1 subunit 1 |
| BLOC1S3 | A0A0D9SE31 | 103234853 | *Chlorocebus sabaeus* | Biogenesis of lysosomal organelles complex 1 subunit 3 |
| BLOC1S3 | Q6QNY0 | 388552 | *Homo sapiens* | Biogenesis of lysosomal organelles complex 1 subunit 3 |
| Bloc1s3 | Q5U5M8 | 232946 | *Mus musculus* | Biogenesis of lysosomal organelles complex 1 subunit 3 |
| BLOC1S3 | A0A218UH38 | | *Lonchura striata* | Biogenesis of lysosomal organelles complex 1 subunit 3 |
| BNC2 | A0A093CJE8 | 104466556 | *Pterocles gutturalis* | Zinc finger protein basonuclin-2 |
| BNC2 | F1NKP6 | 431604 | *Gallus gallus* | Basonuclin 2 |
| BNC2 | H0Z221 | 100221613 | *Taeniopygia guttata* | Basonuclin 2 |
| BRAP | A0A087RAZ7 | 103902848 | *Aptenodytes forsteri* | BRCA1-associated protein |
| BRAP | G1N173 | 100548204 | *Meleagris gallopavo* | BRCA1 associated protein |
| BRAP | Q5ZL09 | 772041 | *Gallus gallus* | Uncharacterized protein |
| BRAP | R0KKU3 | 101803589 | *Anas platyrhynchos* | BRCA1-associated protein |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| BRAT1 | A0A091IAY3 | 103532755 | *Calypte anna* | BRCA1-associated ATM activator 1 |
| BRAT1 | A0A091KWY1 | 104483021 | *Chlamydotis macqueenii* | BRCA1-associated ATM activator 1 |
| BRAT1 | A0A091V949 | 104015091 | *Nipponia nippon* | BRCA1-associated ATM activator 1 |
| BRAT1 | A0A093FXC2 | 104297137 | *Picoides pubescens* | BRCA1-associated ATM activator 1 |
| BRAT1 | G1MVC3 | 100550354 | *Meleagris gallopavo* | BRCA1 associated ATM activator 1 |
| BRAT1 | H0ZCN5 | 100227055 | *Taeniopygia guttata* | BRCA1 associated ATM activator 1 |
| BRAT1 | U3I6B0 | 101805396 | *Anas platyrhynchos* | BRCA1 associated ATM activator 1 |
| BRAT1 | U3KHI3 | 101816597 | *Ficedula albicollis* | BRCA1 associated ATM activator 1 |
| BRAT1L | A0A1D5PJI6 | 107057032 | *Gallus gallus* | Uncharacterized protein |
| BRCA1 | Q90Z51 | 373983 | *Gallus gallus* | Breast and ovarian cancer susceptibility-like protein |
| BRCA1 | U3I714 | 101789679 | *Anas platyrhynchos* | BRCA1, DNA repair associated |
| BRCA1 | U3JDH4 | 101814437 | *Ficedula albicollis* | Uncharacterized protein |
| BRCC3 | E1C8A6 | 422201 | *Gallus gallus* | BRCA1/BRCA2-containing complex subunit 3 |
| BRCC3 | U3I9S5 | 101800040 | *Anas platyrhynchos* | Uncharacterized protein |
| BRCC3 | U3KAZ1 | 101808413 | *Ficedula albicollis* | BRCA1/BRCA2-containing complex subunit 3 |
| BRIP1 | A0A087R033 | 103898034 | *Aptenodytes forsteri* | Fanconi anemia group J protein |
| BRIP1 | A0A091P9C6 | 104318099 | *Haliaeetus albicilla* | Fanconi anemia group J protein |
| BRIP1 | A0A091SND1 | 104028195 | *Pelecanus crispus* | Fanconi anemia group J protein |
| BRIP1 | A0A091UGH3 | 104619663 | *Phaethon lepturus* | Fanconi anemia group J protein |
| BRIP1 | A0A093KLT1 | 104083670 | *Fulmarus glacialis* | Fanconi anemia group J protein |
| BRIP1 | A0A093R258 | 104051424 | *Phalacrocorax carbo* | Fanconi anemia group J protein |
| BRIP1 | A0A0A0ABS1 | 104284355 | *Charadrius vociferus* | Fanconi anemia group J protein |
| BRIP1 | Q3YK19 | 417642 | *Gallus gallus* | Fanconi anemia group J protein homolog |
| BRIP1 | G1N6L5 | 100540611 | *Meleagris gallopavo* | BRCA1 interacting protein C-terminal helicase 1 |
| BRIP1 | R0LTQ8 | 101791710 | *Anas platyrhynchos* | Fanconi anemia group J protein-like protein |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| CORIN | A0A091NM42 | 104276331 | *Apaloderma vittatum* | Atrial natriuretic peptide-converting enzyme |
| CORIN | H0ZD15 | 100231807 | *Taeniopygia guttata* | Uncharacterized protein |
| CORIN | R0L4W7 | 101803071 | *Anas platyrhynchos* | Corin, serine peptidase |
| DNMT1 | Q92072 | 396011 | *Gallus gallus* | DNA (cytosine-5)-methyltransferase 1 |
| DRD2 | A9YZQ5 | 428252 | *Gallus gallus* | Dopamine receptor D2 |
| DRD2 | O73810 | 100303660 | *Meleagris gallopavo* | D(2) dopamine receptor |
| DRD2 | H0YPR8 | 100191007 | *Taeniopygia guttata* | Dopamine receptor D2 |
| DRD2 | U3IND7 | 101804061 | *Anas platyrhynchos* | Dopamine receptor D2 |
| EDA | A0A1D5PK40 | 769069 | *Gallus gallus* | Ectodysplasin A |
| EDA | H0YWU6 | 100220755 | *Taeniopygia guttata* | Ectodysplasin A |
| EDAR | H0ZGY0 | 100221815 | *Taeniopygia guttata* | Ectodysplasin A receptor |
| EDAR | R7VWV2 | 102087618 | *Columba livia* | Tumor necrosis factor receptor superfamily member EDAR |
| EDAR | U3J8N7 | 101798887 | *Anas platyrhynchos* | Ectodysplasin A receptor |
| edn3 | A0A096NNF5 | 101003529 | *Papio anubis* | Endothelin 3 |
| edn3 | A0A0S7LSX7 | | *Poeciliopsis prolifica* | EDN3 |
| edn3 | E0CZ86 | | *Mus musculus* | Endothelin-3 |
| EDNRB | A0A094K7N2 | 104531749 | *Antrostomus carolinensis* | Uncharacterized protein |
| EDNRB | H0ZQR0 | 100228255 | *Taeniopygia guttata* | Endothelin receptor type B |
| EDNRB | U3IXX1 | 101790262 | *Anas platyrhynchos* | Endothelin receptor type B |
| EDNRB2 | F6SB94 | 373909 | *Gallus gallus* | Uncharacterized protein |
| EDNRB2 | F6SBA2 | 373909 | *Gallus gallus* | Uncharacterized protein |
| EDNRB2 | Q8JHV3 | 373909 | *Gallus gallus* | Endothelin receptor type B2 |
| EGFR | P13387 | 396494 | *Gallus gallus* | Epidermal growth factor receptor |
| EGFR | H0YU95 | 100221035 | *Taeniopygia guttata* | Receptor protein-tyrosine kinase |
| EGFR | U3I5P3 | 101797830 | *Anas platyrhynchos* | Receptor protein-tyrosine kinase |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| EGFR | U3JQX4 | 101818290 | *Ficedula albicollis* | Receptor protein-tyrosine kinase |
| ELOVL3 | E7E8G6 | 770955 | *Gallus gallus* | Elongation of very long chain fatty acids protein |
| ELOVL3 | R4GLY7 | 770955 | *Gallus gallus* | Elongation of very long chain fatty acids protein |
| FGFR2 | A0A091Q458 | 104350125 | *Leptosomus discolor* | Fibroblast growth factor receptor |
| FGFR2 | F1NEE9 | 396259 | *Gallus gallus* | Fibroblast growth factor receptor |
| FGFR2 | H0ZLZ4 | 100223379 | *Taeniopygia guttata* | Fibroblast growth factor receptor |
| FIG4 | A0A091NPF1 | 104277194 | *Apaloderma vittatum* | Polyphosphoinositide phosphatase |
| FIG4 | H0ZPC1 | 100229159 | *Taeniopygia guttata* | FIG4 phosphoinositide 5-phosphatase |
| GGT1 | F1NVY4 | 416945 | *Gallus gallus* | Uncharacterized protein |
| GGT1 | H0ZIK1 | 100219009 | *Taeniopygia guttata* | Uncharacterized protein |
| GGT1 | U3IKT8 | 101799199 | *Anas platyrhynchos* | Uncharacterized protein |
| GGT1 | U3JXW3 | 101809751 | *Ficedula albicollis* | Uncharacterized protein |
| GNA11 | P45645 | 100151748 | *Meleagris gallopavo* | Guanine nucleotide-binding protein subunit alpha-11 |
| GNA11 | H0YQK2 | 100219505 | *Taeniopygia guttata* | G protein subunit alpha 11 |
| GNA11 | Q71RI7 | 374077 | *Gallus gallus* | Guanine nucleotide-binding protein G11 alpha-subunit |
| GNA11 | U3I8L6 | 101802177 | *Anas platyrhynchos* | G protein subunit alpha 11 |
| GNAQ | A0A091SUP8 | 104402875 | *Nestor notabilis* | Guanine nucleotide-binding protein G(Q) subunit alpha |
| GNAQ | Q5F3B5 | 427262 | *Gallus gallus* | G protein subunit alpha q |
| GPNMB | F1NPS6 | 428431 | *Gallus gallus* | Glycoprotein nmb |
| GPNMB | G1NDG7 | 100548752 | *Meleagris gallopavo* | Glycoprotein nmb |
| GPNMB | H0YX20 | 100218700 | *Taeniopygia guttata* | Uncharacterized protein |
| GPNMB | Q90372 | 107309209 | *Coturnix japonica* | Protein QNR-71 |
| GPNMB | U3J0D1 | 101795842 | *Anas platyrhynchos* | Glycoprotein nmb |
| HERC2 | E1BW48 | 418682 | *Gallus gallus* | HECT and RLD domain containing E3 ubiquitin protein ligase 2 |
| HERC2 | H0ZJQ4 | 100226646 | *Taeniopygia guttata* | HECT and RLD domain, contains E3 ubiquitin ligase 2 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| HERC2 | U3I746 | 101796116 | *Anas platyrhynchos* | HECT and RLD domain, contains E3 ubiquitin ligase 2 |
| HERC2 | U3JN73 | 101810775 | *Ficedula albicollis* | HECT and RLD domain, contains E3 ubiquitin protein 2 |
| IRF4 | A0A093BMM9 | 104377901 | *Tauraco erythrolophus* | Interferon regulatory factor 4 |
| IRF4 | A0A099ZHV4 | 104573389 | *Tinamus guttatus* | Interferon regulatory factor 4 |
| IRF4 | H0YV95 | 100223845 | *Taeniopygia guttata* | Interferon regulatory factor 4 |
| IRF4 | Q98TX7 | 374179 | *Gallus gallus* | Interferon regulatory factor 4 |
| KITLG | A0A024RBC0 | 4254 | *Homo sapiens* | Kit ligand |
| KITLG | A0A1B2TT44 | 397509 | *Sus scrofa* | Kit ligand |
| Kitlg | A0A250XXE1 | 109685488 | *Castor canadensis* | Kit ligand |
| KITLG | Q2I093 | 100009028 | *Oryctolagus cuniculus* | Kit ligand |
| Kitlg | Q54A14 | 60427 | *Rattus norvegicus* | Kit ligand |
| KRT1 | A0A096NAU0 | 101003537 | *Papio anubis* | Keratin 1 |
| Krt1 | A0A0G2JST3 | 300250 | *Rattus norvegicus* | Keratin, type II cytoskeletal 1 |
| Krt1 | I3M820 | 101971797 | *Ictidomys tridecemlineatus* | Keratin 1 |
| KRT10 | E1C6Q9 | 771977 | *Gallus gallus* | Uncharacterized protein |
| KRT12 | F1NDN6 | 428314 | *Gallus gallus* | Keratin 12 |
| KRT12 | A0A218UN40 | | *Lonchura striata* | Keratin, type I cytoskeletal 12 |
| KRT12 | U3JC60 | | *Ficedula albicollis* | Keratin 12 |
| KRT13 | A0A218UMF6 | | *Lonchura striata* | Keratin, type I cytoskeletal 13 |
| KRT14 | A0A1D5PZ89 | 408039 | *Gallus gallus* | Keratin, type I cytoskeletal 14 |
| KRT14 | Q6PVZ1 | 408039 | *Gallus gallus* | Keratin, type I cytoskeletal 14 |
| KRT14 | A0A218UMW9 | | *Lonchura striata* | Keratin, type I cytoskeletal 14 |
| KRT14 | F1NPR7 | | *Gallus gallus* | Keratin, type I cytoskeletal 14 |
| KRT15 | A0A1D5NZL6 | 408040 | *Gallus gallus* | Uncharacterized protein |
| KRT15 | A0A1D5P5Z4 | 408040 | *Gallus gallus* | Uncharacterized protein |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| KRT15 | F1NDN5 | 408040 | *Gallus gallus* | Uncharacterized protein |
| KRT17 | A0A1D5PXW6 | 100858439 | *Gallus gallus* | Keratin, type I cytoskeletal 14 |
| KRT17 | A0A218ULZ4 | | *Lonchura striata* | Keratin, type I cytoskeletal 17 |
| KRT4 | H0Z0C5 | | *Taeniopygia guttata* | Keratin 4 |
| KRT4 | U3IFQ5 | | *Anas platyrhynchos* | Keratin 4 |
| KRT4 | U3JP95 | | *Ficedula albicollis* | Keratin 4 |
| KRT4 | U3JP96 | | *Ficedula albicollis* | Keratin 4 |
| KRT5 | Q6PVZ5 | 407779 | *Gallus gallus* | Type II alpha-keratin IIA |
| KRT5 | A0A1L1RIW5 | | *Gallus gallus* | Uncharacterized protein |
| KRT7 | O93532 | 395772 | *Gallus gallus* | Keratin, type II cytoskeletal cochleal |
| KRT7 | A0A146F0A0 | | *Gallus gallus* | Keratin, type II cytoskeletal cochleal |
| KRT7 | A0A1L1RNH7 | | *Gallus gallus* | Keratin 7 |
| KRT7 | U3JPW5 | | *Ficedula albicollis* | Keratin 7 |
| KRT8 | Q969I0 | | *Homo sapiens* | KRT8 protein |
| KRT8 | W5ULL9 | | *Ictalurus punctatus* | Keratin, type II cytoskeletal 8 |
| KRT8 | A0A1V4J5G0 | | *Patagioenas fasciata* | Keratin, type II cytoskeletal 8 |
| KRT8 | A0A218UCT8 | | *Lonchura striata* | Keratin, type II cytoskeletal 8 |
| KRT8 | U3JP89 | | *Ficedula albicollis* | Keratin 8 |
| KRT9L | E1C8E4 | 420041 | *Gallus gallus* | Uncharacterized protein |
| LMX1A | A0A091JJY2 | 104132100 | *Egretta garzetta* | LIM homeobox transcription factor 1-alpha |
| LMX1A | E1C2D6 | 777133 | *Gallus gallus* | LIM homeobox transcription factor 1 alpha |
| LMX1A | G1MX42 | 100549066 | *Meleagris gallopavo* | LIM homeobox transcription factor 1 alpha |
| LMX1A | U3IA94 | 101795021 | *Anas platyrhynchos* | LIM homeobox transcription factor 1 alpha |
| LMX1A | U3JZR7 | 101809606 | *Ficedula albicollis* | LIM homeobox transcription factor 1 alpha |
| LOC100220827 | H0Z814 | 100220827 | *Taeniopygia guttata* | Uncharacterized protein |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| LOC100223308 | H0YR95 | 100223308 | *Taeniopygia guttata* | Uncharacterized protein |
| LOC100226448 | H0ZGH0 | 100226448 | *Taeniopygia guttata* | Mitogen-activated protein kinase kinase 1 |
| LOC100229407 | H0YVX6 | 100229407 | *Taeniopygia guttata* | Keratin 12 |
| LOC100230245 | H0YYZ8 | 100230245 | *Taeniopygia guttata* | G protein subunit alpha q |
| LOC100540905 | G1NG34 | 100540905 | *Meleagris gallopavo* | Fibroblast growth factor receptor |
| LOC100540905 | G3UTQ6 | 100540905 | *Meleagris gallopavo* | Fibroblast growth factor receptor 2 |
| LOC100542101 | G1MVH4 | 100542101 | *Meleagris gallopavo* | Keratin 12 |
| LOC100545322 | G1NF62 | 100545322 | *Meleagris gallopavo* | Keratin 7 |
| LOC100546423 | G1N789 | 100546423 | *Meleagris gallopavo* | Uncharacterized protein |
| LOC100547338 | G1N3A2 | 100547338 | *Meleagris gallopavo* | Uncharacterized protein |
| LOC100550119 | G1NPI7 | 100550119 | *Meleagris gallopavo* | HECT and RLD domain, contains E3 ubiquitin ligase 2 |
| LOC100550844 | G1NET1 | 100550844 | *Meleagris gallopavo* | MOS proto-oncogene, serine/threonine kinase |
| LOC101791244 | U3IL97 | 101791244 | *Anas platyrhynchos* | Keratin 12 |
| LOC101794123 | U3IBV8 | 101794123 | *Anas platyrhynchos* | G protein subunit alpha q |
| LOC101797203 | U3I0A3 | 101797203 | *Anas platyrhynchos* | Keratin 7 |
| LOC101801774 | U3I6I4 | 101801774 | *Anas platyrhynchos* | ADAM metallopeptidase with thrombospondin 1M20 |
| LOC101804862 | U3IAT7 | 101804862 | *Anas platyrhynchos* | Uncharacterized protein |
| LOC101805249 | A0A0C5B1C0 | 101805249 | *Anas platyrhynchos* | EDNRB2 protein |
| LOC101809825 | U3JC50 | 101809825 | *Ficedula albicollis* | Uncharacterized protein |
| LOC101815148 | U3K0H5 | 101815148 | *Ficedula albicollis* | Uncharacterized protein |
| LOC104574045 | A0A099YW49 | 104574045 | *Tinamus guttatus* | E3 ubiquitin-protein ligase HERC2 |
| LOC107312830 | A0PAT8 | 107312830 | *Coturnix japonica* | Endothelin receptor B2 |
| LOC107312830 | A1IHM4 | 107312830 | *Coturnix japonica* | Endothelin receptor B2 |
| LYST | A0A094KVS9 | 104528482 | *Antrostomus carolinensis* | Lysosomal-trafficking regulator |
| LYST | A0A1D5PA68 | 421514 | *Gallus gallus* | Lysosomal trafficking regulator |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| LYST | G1NH36 | 100550123 | *Meleagris gallopavo* | Lysosomal trafficking regulator |
| LYST | R0JM22 | 101802189 | *Anas platyrhynchos* | Lysosomal-trafficking regulator |
| LYST | U3KGE0 | 101812819 | *Ficedula albicollis* | Lysosomal trafficking regulator |
| MAP2K1 | A0A091PQD6 | 104322861 | *Haliaeetus albicilla* | Uncharacterized protein |
| MAP2K1 | A0A091S635 | 104408199 | *Nestor notabilis* | Uncharacterized protein |
| MAP2K1 | Q91447 | 103816115 | *Serinus canaria* | Dual specificity mitogen-activated protein kinase 1 |
| MAP2K1 | Q5ZIF0 | 415549 | *Gallus gallus* | Mitogen-activated protein kinase kinase 1 |
| MAP2K1 | U3J7A4 | 101791037 | *Anas platyrhynchos* | Mitogen-activated protein kinase kinase 1 |
| MC1R | H0ZCF4 | 100225710 | *Taeniopygia guttata* | Melanocyte-stimulating hormone receptor |
| MC1R | Q59I39 | 427562 | *Gallus gallus* | Melanocyte-stimulating hormone receptor |
| MC1R | T2AWX2 | 106145744 | *Columba livia* | Melanocyte-stimulating hormone receptor |
| MCOLN3 | F1NIW3 | 429091 | *Gallus gallus* | Mucolipin 3 |
| MCOLN3 | H0Z7X7 | 100218399 | *Taeniopygia guttata* | Uncharacterized protein |
| MCOLN3 | U3JUJ5 | 101809471 | *Ficedula albicollis* | Mucolipin 3 |
| MGRN1 | A0A1D5PJI2 | 416660 | *Gallus gallus* | Mahogunin ring finger 1 |
| MGRN1 | U3IMT8 | 101801804 | *Anas platyrhynchos* | Mahogunin ring finger 1 |
| MGRN1 | U3JNR2 | 101815667 | *Ficedula albicollis* | Mahogunin ring finger 1 |
| MITF | A0A091MWY0 | 104273398 | *Apaloderma vittatum* | Microphthalmia-associated transcription factor |
| MITF | A0A094L060 | 104529677 | *Antrostomus carolinensis* | Microphthalmia-associated transcription factor |
| MITF | H0ZHV0 | 100232286 | *Taeniopygia guttata* | Melanogenesis associated transcription factor |
| MITF | Q9IAU0 | 395886 | *Gallus gallus* | Microphthalmia protein |
| MLPH | B0F2C1 | 424019 | *Gallus gallus* | Melanophilin transcript variant 1 |
| MLPH | C0ITK8 | 102093006 | *Columba livia* | Melanophilin transcript variant 1 |
| MLPH | U3IBW0 | 101793176 | *Anas platyrhynchos* | Melanophilin |
| OSTM1 | A0A093BLD0 | 104388159 | *Chaetura pelagica* | Osteopetrosis-associated transmembrane protein 1 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| OSTM1 | Q5ZMW4 | 421773 | *Gallus gallus* | Osteopetrosis associated transmembrane protein 1 |
| PAX3 | H0ZBP2 | 100222498 | *Taeniopygia guttata* | Paired box 3 |
| PAX3 | I7EXJ2 | 101797445 | *Anas platyrhynchos* | Paired box 3 |
| PAX3 | Q8QGS3 | 374127 | *Gallus gallus* | Paired-box transcription factor protein PAX3 |
| Pomc | P01193 | 18976 | *Mus musculus* | Met-enkephalin |
| Pomc | Q8K422 | 24664 | *Rattus norvegicus* | Proopiomelanocortin |
| PRKAR2B | A0A093ISZ2 | 104507587 | *Eurypyga helias* | cAMP-dependent protein kinase type II-beta regulatory subunit |
| PRKAR2B | F1NC65 | 769420 | *Gallus gallus* | Protein kinase cAMP-dependent type II regulatory |
| PRKAR2B | G1N4X2 | 100547863 | *Meleagris gallopavo* | Protein kinase cAMP-dependent type II regulatory |
| RECQL4 | A0A0D9R7C3 | 103237633 | *Chlorocebus sabaeus* | RecQ like helicase 4 |
| Recql4 | A0A1S3GKJ1 | 105999007 | *Dipodomys ordii* | ATP-dependent DNA helicase Q4 isoform X2 |
| RPL24 | B5FYC4 | 100190138 | *Taeniopygia guttata* | Putative ribosomal protein L24 variant 2 |
| RPL24 | E1C8F7 | 418401 | *Gallus gallus* | Ribosomal protein L24 |
| RPL24 | U3I7V0 | 101793687 | *Anas platyrhynchos* | Ribosomal protein L24 |
| RPS19 | A0A1D5PDV6 | 107050719 | *Gallus gallus* | Ribosomal protein S19 |
| RPS20 | A0A091RA86 | 103780004 | *Merops nubicus* | 40S ribosomal protein S20 |
| RPS20 | F1NH93 | 430990 | *Gallus gallus* | Ribosomal protein S20 |
| RPS20 | H0ZLG5 | 100227916 | *Taeniopygia guttata* | Ribosomal protein S20 |
| SFXN1 | A0A1D5PGI3 | 416221 | *Gallus gallus* | Sideroflexin |
| SFXN1 | H0YPZ8 | 100226188 | *Taeniopygia guttata* | Sideroflexin |
| SFXN1 | U3INW1 | 101798095 | *Anas platyrhynchos* | Sideroflexin |
| SLC24A4 | E1BZM8 | 772279 | *Gallus gallus* | Solute carrier family 24 member 4 |
| SLC24A4 | G1NKC1 | 100540694 | *Meleagris gallopavo* | Solute carrier family 24 member 4 |
| SLC24A4 | U3IH87 | 101789854 | *Anas platyrhynchos* | Solute carrier family 24 member 4 |
| SLC36A1 | A0A1D5NUH4 | 770250 | *Gallus gallus* | Uncharacterized protein |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| SLC45A2 | A0A093BZV9 | 104472133 | *Pterocles gutturalis* | Membrane-associated transporter protein |
| SLC45A2 | A0A093GN13 | 104368370 | *Tyto alba* | Membrane-associated transporter protein |
| SLC45A2 | A0A093NSZ4 | 103912208 | *Pygoscelis adeliae* | Membrane-associated transporter protein |
| SLC45A2 | H0YUN5 | 100228141 | *Taeniopygia guttata* | Solute carrier family 45 member 2 |
| SLC7A11 | A0A093TLJ2 | 104051330 | *Phalacrocorax carbo* | Cystine/glutamate transporter |
| SLC7A11 | E1C734 | 428731 | *Gallus gallus* | Solute carrier family 7 member 11 |
| SLC7A11 | G1MUY1 | 100541353 | *Meleagris gallopavo* | Solute carrier family 7 member 11 |
| SLC7A11 | U3J347 | 101797014 | *Anas platyrhynchos* | Solute carrier family 7 member 11 |
| SLC7A11 | U3JQL4 | 101809120 | *Ficedula albicollis* | Solute carrier family 7 member 11 |
| SMARCA5 | E1C0M8 | 422457 | *Gallus gallus* | SWI/SNF related, actin dependent regulator of chromatin |
| SMARCA5 | H0YW21 | 100223859 | *Taeniopygia guttata* | SWI/SNF related, actin dependent regulator of chromatin |
| SMARCA5 | U3IP13 | 101793520 | *Anas platyrhynchos* | SWI/SNF related, actin dependent regulator of chromatin |
| SNAI2 | A0A094K3V3 | 104532646 | *Antrostomus carolinensis* | Zinc finger protein SNAI2 |
| SNAI2 | H0ZKY0 | 100231698 | *Taeniopygia guttata* | Snail family transcriptional repressor 2 |
| SNAI2 | R0KCR9 | 101804065 | *Anas platyrhynchos* | Zinc finger protein SNAI2 |
| SNAI2L | A0A1D5PN27 | 107055962 | *Gallus gallus* | Uncharacterized protein |
| SOX10 | A0A094L837 | 104526574 | *Antrostomus carolinensis* | Transcription factor SOX-10 |
| SOX10 | H0ZJN2 | 100222412 | *Taeniopygia guttata* | SRY-box 10 |
| SOX10 | R9PXP0 | 395573 | *Gallus gallus* | Transcription factor SOX-10 |
| SOX18 | H0ZA76 | 100224917 | *Taeniopygia guttata* | SRY-box 18 |
| SOX18 | Q90ZA9 | 374200 | *Gallus gallus* | Transcription factor SOX18 |
| STX17 | A0A091I2H8 | 103530691 | *Calypte anna* | Syntaxin-17 |
| STX17 | A0A093PUL7 | 103756784 | *Manacus vitellinus* | Syntaxin-17 |

Table continued.

| Symbol | Accession | Entrez_ID | Taxon_Name | Protein_Name |
|---|---|---|---|---|
| STX17 | H0ZEU8 | 100223398 | *Taeniopygia guttata* | Syntaxin 17 |
| TBX15 | G1NN54 | 100541588 | *Meleagris gallopavo* | T-box 15 |
| TBX15 | R0JGZ7 | 101792140 | *Anas platyrhynchos* | T-box transcription factor TBX15 |
| TBX15L | A0A1D5PDD9 | 100858059 | *Gallus gallus* | Uncharacterized protein |
| TFAP2A | O13111 | 395982 | *Gallus gallus* | AP-2 transcription factor |
| TPCN2 | F1NSV9 | 423141 | *Gallus gallus* | Two pore segment channel 2 |
| VPS33A | A0A091HL02 | 103528310 | *Calypte anna* | Vacuolar protein sorting-associated protein 33A |
| VPS33A | A0A091L9H1 | 104475817 | *Chlamydotis macqueenii* | Vacuolar protein sorting-associated protein 33A |
| VPS33A | H0Z751 | 100220818 | *Taeniopygia guttata* | Uncharacterized protein |
| WNT1 | Q91029 | 396160 | *Gallus gallus* | Protein Wnt-1 |
| WNT3A | A0A1D5PX88 | 395396 | *Gallus gallus* | Protein Wnt |
| WNT3A | R0LD75 | 101800064 | *Anas platyrhynchos* | Protein Wnt |
| WNT3A | Q2LMP1 | 395396 | *Gallus gallus* | Protein Wnt-3a |
| ZIC2 | A0A1D5PNB7 | 428021 | *Gallus gallus* | Zic family member 2 |

Table S4.7. Transcripts with hits to *BLOC*-complex genes and gene products in a blastx search against the nr database.

| Sequence Name | Description | Sequence Length | E-Value | Blast Similarity | Blast Hit Taxon |
|---|---|---|---|---|---|
| TRINITY_DN35505_c0_g1_i1 | *BLOC-1S1* | 450 | 3.40E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g2_i1 | *BLOC-1S1* | 449 | 3.20E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g3_i1 | *BLOC-1S1* | 417 | 4.60E-59 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g4_i1 | *BLOC-1S1* | 450 | 3.40E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g5_i1 | *BLOC-1S1* | 450 | 4.60E-50 | 98% | *Latimeria chalumnae* |
| TRINITY_DN35505_c0_g6_i1 | *BLOC-1S1* | 450 | 4.60E-50 | 98% | *Latimeria chalumnae* |
| TRINITY_DN35505_c0_g7_i1 | *BLOC-1S1* | 450 | 3.40E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g8_i1 | *BLOC-1S1* | 384 | 1.40E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g9_i1 | *BLOC-1S1* | 450 | 3.40E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g10_i1 | *BLOC-1S1* | 449 | 3.00E-50 | 98% | *Latimeria chalumnae* |
| TRINITY_DN35505_c0_g11_i1 | *BLOC-1S1* | 450 | 3.40E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN35505_c0_g12_i1 | *BLOC-1S1* | 450 | 3.40E-54 | 97% | *Patagioenas fasciata monilis* |
| TRINITY_DN47443_c0_g1_i1 | *BLOC-1S2* | 433 | 3.40E-44 | 96% | *Anser cygnoides domesticus* |
| TRINITY_DN47443_c0_g1_i3 | *BLOC-1S2* | 1080 | 2.40E-67 | 100% | *Calypte anna* |
| TRINITY_DN47443_c0_g1_i2 | *BLOC-1S2*, partial | 1247 | 4.90E-39 | 100% | *Calypte anna* |
| TRINITY_DN49905_c0_g1_i1 | *BLOC-1S4* | 1502 | 1.10E-91 | 92% | *Haliaeetus leucocephalus* |
| TRINITY_DN46750_c0_g1_i1 | *BLOC-1S5*, partial | 894 | 5.00E-101 | 99% | *Calypte anna* |
| TRINITY_DN46058_c0_g1_i1 | *BLOC-1S6* | 1020 | 1.90E-61 | 100% | *Calypte anna* |
| TRINITY_DN46058_c0_g1_i2 | *BLOC-1S6*, partial | 462 | 1.10E-18 | 100% | *Calypte anna* |

Table S4.8. Transcripts with hits to BLOC-complex genes and gene products in a blastx search against the pigmentation gene database.

| Sequence Name | Sequence Description | Sequence Length | E-Value | Blast Similarity |
|---|---|---|---|---|
| TRINITY_DN35505_c0_g5_i1 | *BLOC-1S1* | 450 | 6.80E-54 | 96% |
| TRINITY_DN35505_c0_g6_i1 | *BLOC-1S1* | 450 | 6.80E-54 | 96% |
| TRINITY_DN35505_c0_g1_i1 | *BLOC-1S1* | 450 | 2.40E-54 | 94% |
| TRINITY_DN35505_c0_g4_i1 | *BLOC-1S1* | 450 | 2.40E-54 | 94% |
| TRINITY_DN35505_c0_g7_i1 | *BLOC-1S1* | 450 | 2.40E-54 | 94% |
| TRINITY_DN35505_c0_g9_i1 | *BLOC-1S1* | 450 | 2.40E-54 | 94% |
| TRINITY_DN35505_c0_g10_i1 | *BLOC-1S1* | 449 | 1.20E-53 | 96% |
| TRINITY_DN35505_c0_g11_i1 | *BLOC-1S1* | 450 | 2.40E-54 | 94% |
| TRINITY_DN35505_c0_g12_i1 | *BLOC-1S1* | 450 | 2.40E-54 | 94% |
| TRINITY_DN35505_c0_g2_i1 | *BLOC-1S1* | 449 | 2.30E-54 | 94% |
| TRINITY_DN35505_c0_g8_i1 | *BLOC-1S1* | 384 | 1.00E-54 | 94% |
| TRINITY_DN35505_c0_g3_i1 | *BLOC-1S1* | 417 | 2.60E-55 | 94% |
| TRINITY_DN47443_c0_g1_i2 | *BLOC-1S2* | 1247 | 7.80E-45 | 98% |
| TRINITY_DN47443_c0_g1_i1 | *BLOC-1S3* | 433 | 4.40E-49 | 98% |
| TRINITY_DN47443_c0_g1_i3 | *BLOC-1S4* | 1080 | 3.00E-58 | 97% |
| TRINITY_DN49905_c0_g1_i1 | *BLOC-1S4* | 1502 | 1.50E-94 | 89% |
| TRINITY_DN52506_c0_g1_i6 | *BLOC-1S4* | 4377 | 8.60E-06 | 45% |
| TRINITY_DN52506_c0_g1_i2 | *BLOC-1S4* | 4282 | 8.40E-06 | 45% |
| TRINITY_DN46058_c0_g1_i1 | *BLOC-1S6* | 1020 | 8.80E-59 | 97% |
| TRINITY_DN46058_c0_g1_i2 | *BLOC-1S6* | 462 | 5.80E-19 | 92% |

## APPENDIX C.

## PUBLISHING AGREEMENT

MPublishing Model Journal Publishing Agreement
Author Copyright

The following is an agreement between

Caroline Judy _____ (the "Corresponding Author")

acting on behalf of all authors of the work ("Authors") and

The Journal of Caribbean Ornithology _____(the "Journal")

which governs

Density and Abundance of the Black-billed Streamertail (Trochilus s _____ (the "Work").

Whereas the parties desire to promote effective scholarly communication that promotes local control of intellectual assets, the parties for valuable consideration agree as follows.

A. **CORRESPONDING AUTHOR'S GRANT OF RIGHTS**
The Corresponding Author grants to the Journal, during the full term of copyright and any extensions or renewals of that term, the following:

1. An irrevocable non-exclusive right to reproduce, republish, transmit, sell, distribute, and otherwise use the Work in electronic and print editions of the Journal and in derivative works throughout the world, in all languages, and in all media now known or later developed.
2. An irrevocable non-exclusive right to create and store electronic archival copies of the Work, including the right to deposit the Work in open access digital repositories.
3. An irrevocable non-exclusive right to license others to reproduce, republish, transmit, and distribute the Work under the condition that the Authors are attributed. (Currently this is carried out by publishing the content under a Creative Commons Attribution 3.0 license.)

Copyright in the Work remains with the Authors.

B. **CORRESPONDING AUTHOR'S DUTIES**
1. When distributing or re-publishing the Work, the Corresponding Author agrees to credit the Journal as the place of first publication.
2. The Corresponding Author agrees to inform the Journal of any changes in contact information.

C. **CORRESPONDING AUTHOR'S WARRANTY**
The Corresponding Author represents and warrants that the Work is the Authors' original work and that it does not violate or infringe the law or the rights of any third party and, specifically, that the Work contains no matter that is defamatory or that infringes any

literary or proprietary rights, intellectual property rights, or any rights of privacy. The Corresponding Author also warrants that he or she has the full power to make this agreement, and if the Work was prepared jointly, the Corresponding Author agrees to inform the Authors of the terms of this Agreement and to obtain their written permission to sign on their behalf. The Corresponding Author agrees to hold the Journal harmless from any breach of the aforestated representations.

D. **JOURNAL'S DUTIES**
In consideration of the Author's grant of rights, the Journal agrees to publish the Work, attributing the Work to the Authors.

E. **ENTIRE AGREEMENT**
This agreement reflects the entire understanding of the parties. This agreement may be amended only in writing by an addendum signed by the parties. Amendments are incorporated by reference to this agreement.

ACCEPTED AND AGREED BY THE CORRESPONDING AUTHOR ON BEHALF OF ALL AUTHORS CONTRIBUTING TO THIS WORK:

Corresponding Author: Cindira July Date: Nov. 1, 2018

Journal Representative: Jason M. Townsend Date: Oct. 24, 2018

VITA

Caroline Duffie Judy was born in Atlanta, Ga. to Carol and Traywick Duffie. She has an older sister, Lauren, and a twin sister, Amanda. While raised in a city, Caroline was always drawn to the natural world. She spent most of her childhood outside hunting salamandars, frogs, and turtles. Weekend trips to her father's hunting camp deepened her knowledge of the Piedmont forest. Caroline majored in Ecology at the University of Georgia. Her undergraduate advisor, Dr. James Richardson, encouraged her to participate in a Maymester course in Costa Rica, where she first fell in love with the tropics. She spent her senior year in Ecuador refining her Spanish and conducting independent research on Amazonian parrots. After graduating, she worked for two years as a laboratory technician at the Southeastern Cooperative of Wildlife Disease Study in Athens, Ga., before pursuing her Master's degree at University of Missouri-St. Louis under the advisement of Dr. Patricia Parker. She conducted research in the Gálapagos Islands on Flightless Cormorants. Caroline then moved to Washington, D.C. to work at the National Science Foundation, where she broadened her experience with science policy and saved money to return to graduate school. Two years later, she joined Robb Brumfield's lab at Louisiana State University as a PhD student, and conducted research in Jamaica on Streamertail Hummingbirds. She married Dan Judy, a fellow Georgian, and they began their life together. They welcomed a daughter, Rosie, two years later. Caroline and Dan are active members at a Presbyterian church in Alexandria, VA. Caroline enjoys engaging in discussions of science and faith, and sees herself as a bridge builder between these communities. In her free time, Caroline enjoys playing classical guitar, cooking, and spending time out-of-doors with her family.